SPAWAR
Systems Center
San Diego

TECHNICAL REPORT 1816
February 2000

Online Monitoring of Oils in Wastewater Using Combined Ultraviolet Fluorescence and Light Scattering with an Artificial Neural Network

L. M. He
San Diego State
University Foundation

L. L. Kear-Padilla
CSC

S. H. Lieberman
J. M. Andrews
SSC San Diego

SSC San Diego

TECHNICAL REPORT 1816
February 2000

Online Monitoring of Oils in Wastewater Using Combined Ultraviolet Fluorescence and Light Scattering with an Artificial Neural Network

L. M. He
San Diego State
University Foundation

L. L. Kear-Padilla
CSC

S. H. Lieberman
J. M. Andrews
SSC San Diego

SSC San Diego
San Diego, CA 92152–5001

SSC SAN DIEGO
San Diego, California 92152–5001

E. L. Valdes, CAPT, USN
Commanding Officer

R. C. Kolb
Executive Director

ADMINISTRATIVE INFORMATION

The work described in this report was performed jointly for the Office of Naval Research Environmental Quality Technology Program and the Naval Submarine Surface Warfare Center, Carderock Division by the San Diego State University Foundation, Computer Sciences Corporation, and the Environmental Chemistry/Biotechnology Branch (D361), SSC San Diego.

Released by
M. D. Putnam, Head
Environmental Chemistry/
Biotechnology Branch

Under authority of
R. H. Moore, Head
Marine Navigation
Division

This technology may be the subject of one or more invention disclosures assignable to the U.S. Government. Licensing inquiries may be directed to:

Harvey Fendelman
Legal Counsel for Patents
SSC San Diego, D0012
San Diego, CA 92152–5765
(619) 553–3001

SB

EXECUTIVE SUMMARY

Ultraviolet (UV) fluorescence and light scattering are two analytical methods commonly used in instrumentation for online measurement of oils in water. UV fluorescence-based instruments detect both dissolved and emulsified aromatic constituents of oils. Light-scattering-based sensors measure optical scattering induced by emulsified oil droplets. A major technical challenge for each method is to maintain quantitative accuracy in the presence of chemical and physical interferences, including fluorescent organic compounds (e.g., detergents and natural organic matter), suspended solid particles, dissolved salts, etc. To address this issue, we have been developing a new monitoring system that simultaneously combines both UV fluorescence and light scattering spectroscopy. Four major types of oils (lube oils 2190 and 9250, diesel fuel marine [DFM], and JP5), each of which had a dozen subtypes of oil samples, were examined to obtain the intensity of both fluorescence and scattering as a function of oil, detergent (Mil-D and Tide®), and seawater concentrations. Both fluorescence and light scattering intensities varied significantly with oil types and subtypes. Both Mil-D and Tide® greatly influenced the fluorescence and scattering of oil samples.

The tremendous variations in fluorescence and scattering intensity with oil types and subtypes, detergents, and seawater make it difficult to calibrate the analytical instrument using traditional methods; hence, we have implemented a multivariate, nonlinear calibration of instrumental response through an artificial neural network. We have demonstrated that the simultaneous, combined use of fluorescence and scattering data significantly improves quantitative prediction accuracy. The trained backpropagation neural network was used successfully to predict concentrations of single oils and their mixtures, even in the presence of detergents and seawater, and appears well suited for calibrations of an online oil content monitor. The trained network processes information very quickly and is appropriate for real-time applications. The newly developed technique permits the online monitoring of oil spills, the accurate determination of oil concentrations in wastewater discharged from ships.

CONTENTS

Figures

Tables

INTRODUCTION

Online measurement of small quantities of oil in water (e.g., low mg L^{-1}) is extremely difficult in the presence of many physical and chemical interferences. Both fluorescence and light scattering may be used for real-time measurement of oil content in wastewater (Nardella, Raw, and Stokes, 1989; Parker and Pitt, 1986). The fluorescent technique detects the intensity of fluorescence emission from both dissolved and emulsified polycyclic aromatic hydrocarbons (PAHs) when irradiated with ultraviolet light (Andrews and Lieberman, 1998; Lieberman, 1998). This technique is very sensitive to PAH concentration, but the response is oil-type-dependent since oils have varying constituents and PAH content. Constantly revised calibrations of instruments are therefore necessary to allow for oil type variations. In many cases, the instrumental calibration is impossible because the oil composition of a sample (e.g., wastewater) is rarely known. Wastewater discharged from ships and tankers may contain other contaminants such as solid particles and detergents, which significantly influence the online fluorescence measurement of samples, causing even more severe calibration difficulties.

The light scattering methods measure the attenuation of the intensity of light passing through a sample or the intensity of light that is scattered by the sample (Mowery et al., 1997; Nardella, Raw, and Stokes, 1989). One shortcoming of this technique is that it cannot detect the dissolved phase of oil constituents. In addition, the scattering technique poses difficulties in distinguishing between oil droplets and solid particles. Multi-angle scattering and re-emulsification methods reduce the effect induced by particles, but complete compensation for solids content is difficult. The multi-angle scattering correction is based on assumptions that the particle size of solids differs from that of oil droplets, and presumably the former is larger than the latter (Parker and Pitt, 1986). The re-emulsification correction uses ultrasonic techniques to mix oil–water samples. During this mixing process, it is theorized that the oil droplet size is reduced, but solid particles remain the same (Mowery et al., 1997; Parker and Pitt, 1986). However, many rust particles in wastewater will break down during sonication, resulting in analytical errors.

The major technical challenge for each of these methods is to maintain quantitative accuracy for online measuring of oil-in-water concentrations in the presence of unknown chemical and physical interferences, including fluorescent organic compounds, detergents, suspended solid particles, and dissolved salts. Traditional methods for calibrating analytical instruments are clearly not applicable to this complex system, while an artificial neural network (ANN) may be suitable to associate variously derived spectral signals with specific concentrations of oils having various interfering factors.

Artificial neural networks learn by observation, i.e., learn to recognize key features in a data set by repetitiously examining examples of the same or similar data (Zupan and Gasteiger, 1991). This attribute effectively identifies patterns in new or intricate data. There are numerous ANN paradigms, some more effective than others at addressing certain specific processing needs. However, the multilayer backpropagation paradigm is the most widely used (Andrews and Lieberman, 1994). The success and popularity of the backpropagation algorithm are based largely on its proven adeptness at solving pattern recognition problems. Any arbitrarily complex nonlinear associative mapping may be encoded into the weight space or memory of a backpropagation neural network. For example, the backpropagation neural network identifies seven types of oils based on their fluorescence spectra, and this data processing is suitable for real-time discrimination of fuels and oils under various conditions (Andrews and Lieberman, 1994).

This report presents the combined use of fluorescence and scattering spectroscopy for quantitatively predicting single oils and mixtures in wastewater in the presence of interfering substances. A multivariate, nonlinear calibration of instrumental responses is implemented through the multilayer feed-forward backpropagation neural network. This newly developed technique permits accurate online monitoring of oils in natural waters resulting from spillage and leaking, and in wastewater discharging from ships and oil refinery industries.

MATERIALS AND METHODS

EXPERIMENTAL SECTION

Samples

Single and mixed oil samples were prepared with four types of oils: diesel fuel marine (DFM), jet fuel (JP5), and lube oils 2190 and 9250, all of which are commonly used in U.S. Navy ships. The oil concentration ranged from 10 to 120 mg L^{-1} for the training data set, and from 5 to 110 mg L^{-1} for the test data set. Selection of the oil concentration range was based on the current discharge limits set by Federal regulation: 15 mg L^{-1} of oil in the effluent in port and 100 mg L^{-1} at sea (Mowery et al., 1997; Nardella et al., 1989). The serial addition method was used to increase oil concentration in a set of samples. After deionized water was analyzed as a blank, oil was added to the water to obtain a desired concentration, and then the sample was mixed at a low speed with a laboratory blender (stainless steel vessel) for 2 minutes.

Mixed oil samples included binary and quadrate mixtures. The ratios of the mixtures ranged from 1:1 to 1:5. Two detergents (Liquid Tide® and Mil-D), which are commonly used in U.S. Navy ships, and natural seawater were used as interferences. The concentrations of interferences in oil samples were 10, 30, and 60 mg L^{-1} for detergent, and 10, 50, and 100% for seawater.

Fluorescence Measurement

Samples with various oil concentrations were analyzed using the Hitachi model F-4500 fluorescence spectrophotometer (Hitachi Instruments, Inc., Naperville, IL). The fluorimeter was equipped with a 150-W xenon discharge lamp, a reference phototube, an excitation monochromator, an emission monochromator, and a photomultiplier. Fluorescence emission spectra (250 to 550 nm, 2-nm intervals) were collected at five excitation wavelengths (254, 266, 278, 290, and 302 nm). Samples were scanned with 3-mL quartz flow cells having a path length of 1 cm. All spectra were corrected for variations in system response as described in the Hitachi manual.

After mixing, the sample was pumped through the spectrofluorimeter flow cells for 1 minute and scanned. More oil was added, the sample was mixed and analyzed, and the process was repeated until the maximum desired concentration was reached. After each series of measurements, the system was cleaned using hot, soapy water, followed by a hot water rinse and a room-temperature deionized water rinse.

Light intensity in a fluorimeter may change with time because of bulb aging, which affects the intensity of oil fluorescence spectra. To monitor the power intensity of the light source, a standard fluorescing compound, quinine sulfate, was used as a quality control to examine the change in power intensity. The spectrum of quinine sulfate was collected at the beginning and end of each day of measurement. The sum of the spectral intensities of quinine sulfate from 400 to 520 nm was compared to determine if there was a significant change in light source. The data collected for oils were adjusted to a standard status using the intensity of the quinine sulfate spectra whenever necessary.

Light Scattering Measurement

Laser light scattering measurements were performed for each sample using an LISST-100 particle size analyzer (Sequoia Scientific, Inc., Mercer Island, WA). The instrument is composed of a collimated laser beam at 670-nm wavelength, beam manipulation and orienting mounts, a scattered-light receiving lens, and a 32-ring detector. The analyzer first detects the intensity of small-angle scattering induced by oil droplets in water. The scattering data are then inverted to produce the droplet size distribution (i.e., number, area, or volume of droplets per unit volume of water). The instrument outputs 32 size ranges from 1.25 to 250 μ in diameter, with scattering angles detected from 0.0017 to 0.34 radians. The recommended optical transmission is 30 to 99%. When the raw transmitted laser power falls below a factory-set minimum (i.e., optical transmission is very low for a sample of high oil concentration) the processed data are all zeros, which is hereafter referred to as the zero volume concentration distribution.

During analysis, the sample was pumped through the scatter chamber. Once the chamber was filled, the LISST was shaken to release any large bubbles remaining in the chamber. One minute later, the scatter data were obtained and stored in the computer.

ARTIFICIAL NEURAL NETWORK

The neural network used in this study was a fully connected, feed-forward system trained with a backpropagation algorithm. The software was NeuralWorks Professional II Plus (NeuralWare, Inc., Pittsburgh, PA). All the neural network computations were performed on a Pentium® II 450-MHz PC with 128-MB RAM.

The backpropagation neural network (figure 1) constitutes three layers: input, hidden, and output, each layer having varying neurons. The theory for backpropagation is briefly described below. For a more explicit description, see Rumelhart et al. (1986), Wasserman (1989), and Zupan and Gasteiger (1991). Training the backpropagation network requires the following four steps: (1) select a training pair from the training set and apply the input vector to the network input, (2) calculate the output of the network, (3) calculate the error between the network output and the desired output, and (4) adjust the weights of the network in a way that minimizes the error. Steps 1 through 4 for each training pair in the training set are repeated until the error for the entire training set is acceptably low. In step 2, the output calculation is performed on a layer-by-layer basis. Referring to figure 1, first the outputs of the neurons in layer *i* are calculated and the outputs are then used as inputs to layer *j*. The layer *j* outputs are calculated and served as layer *k* inputs. The calculated layer *k* outputs based on the layer *j* outputs are the network outputs.

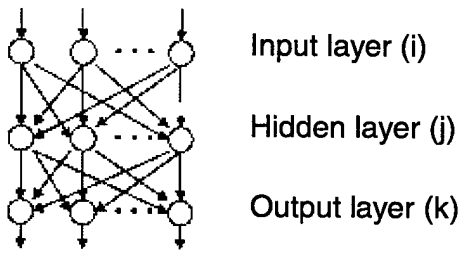

Figure 1. Three-layer backpropagation artificial neural network.

In step 3, each network output is subtracted from its corresponding component of desired output vector to produce an error. This error is used in step 4 to adjust the weights of the network. After enough repetitions of these four steps, the error between actual outputs and desired outputs should be reduced to an acceptable value, and the network is trained. At this point, the network can be used for validation, and weights are not changed. In the validation process, a test pair from a test data set is input to the network, and the output is calculated. A comparison of the actual output with the desired output for the test pair shows how well the network has been trained.

As shown above, training the backpropagation network involves two passes: forward and reverse. In the forward pass (steps 1 and 2), the input signal propagates from the network input to its output. In the reverse pass (steps 3 and 4), the calculated error signal propagates backward through the network.

In the forward pass, the output vector (O_i) from the input layer (i) is given by

$$O_i = f\left(\sum_i I_i W_{ij} + b_j\right), \tag{1}$$

where f is the transfer function, I_i is the input vector in the input layer i, W_{ij} is the connection weight vector between layers i and j, and b_j is the bias term of layer j responsible for accommodating nonzero offsets in the data. The output vector for one layer is the input vector for the next, so calculating the outputs of the final layer requires the application of equation 1 to each layer, from the network's input to its output.

The transfer function used in this study is sigmoid and has the following form:

$$f(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

The sigmoid function compresses the range of summation so that the actual output (O_i) lies between 0 and 1. However, when a MinMax table is used together with the sigmoid function, all input values are mapped to lie between -1.0 and 1.0. A MinMax table was always selected in this study. To improve the network's convergence, all the input values between –1.0 and 1.0 are rescaled to lie in the range from 0.2 to 0.8.

In the reverse pass, all the weights are adjusted starting from the output layer to the hidden layer. Since a desired value is available in each neuron in the output layer, adjustment of the weights associated with the output layer k is readily accomplished using the generalized delta (δ) rule. The δ error (δ_k) for each neuron in the output layer k for each training pair may be expressed as

$$\delta_k = O_k (1 - O_k) (D_k - O_k), \tag{3}$$

where O_k and D_k are the network's actual and desired outputs, respectively, for the output layer.

Since the hidden layer does not have the desired outputs, training is more complicated. The δ error (δ_j) for a neuron in the hidden layer j for each training pair may be expressed as

$$\delta_j = O_j (1 - O_j) \Sigma(\delta_k W_{jk}), \tag{4}$$

where O_j is the output vector in the hidden layer j and W_{jk} is the connection weight vector between the hidden layer j and output layer k.

These δ error values from equations 3 and 4 are used to adjust connection weights:

$$\Delta W(n) = \eta \delta O \tag{5}$$

$$\Delta W(n+1) = \eta \delta O + \alpha[\Delta W(n)] \tag{6}$$

$$W(n+1) = W(n) + \Delta W(n+1) \tag{7}$$

where ΔW is the change in value for the connection weight between two layers, (n) is at step n (before adjustment), $(n + 1)$ is at step $(n + 1)$ (after adjustment), η is the learning rate, α is the momentum term, O is the output value, and W is the connection weight. The momentum term is used to filter out high-frequency fluctuations of the network and to prevent to some degree the convergence process from being trapped into the local minima. Equations 5 through 7 are used for both hidden and output layers. For each neuron in a given layer, δ errors are calculated, and all weights associated with that layer are adjusted. This process is repeated, moving back toward the input layer, layer by layer, until all weights are adjusted. The training process is accomplished when the error for the entire training set is acceptably low.

A critical step in training a network is to optimize a number of parameters used in the network. The parameters include the number of hidden layers, the number of neurons in each hidden layer, the number of inputs and outputs, the transfer functions used in each hidden and output layer, learning rate, and momentum term. Since there are no formal rules to determine these parameters, their selection is based largely on trial and error. Performance of a network is evaluated on its ability to generalize, that is, to what degree a trained network can be used to recognize or predict unknown samples. The error is measured as the root mean square (RMS) at the output layer (Andrews and Lieberman, 1994) as follows:

$$RMS = \sqrt{\frac{\sum_{i}^{n} (O_i - D_i)^2}{n}}, \tag{8}$$

where O_i = actual output at neuron I, D_i = desired output at neuron i, n = the number of training or test pairs used in training or test. A lower RMS indicates a better-trained network, but a trained network with a low RMS may have been over-trained (see below). The best-trained neural network is therefore selected based on prediction percentage. The prediction percentage is calculated by dividing the number of correctly predicted test cases by the total number of cases in a set of test data. For this study, an oil concentration is correctly predicted when the predicted value falls within ±5 mg L^{-1} or ±20%, whichever is greater (Parker and Pitt, 1986).

Since a simple network with a minimum of hidden layers is preferred, the network used throughout this report has only one hidden layer. The general delta rule was employed as the learning rule throughout the computation. A constant momentum of 0.4 was used for training all the networks. Random values between -0.1 and 0.1, and -0.2 and 0.2 were used as the initial weights for the hidden layer and output layer, respectively.

LINEAR MODEL

Linear regression, which is the traditional method for calibrating an analytical instrument, was used as a benchmark for comparing ANN performance results. Both fluorescence and light scattering

may be used to relate the oil concentration to signal intensity. For fluorescence, all intensities over wavelengths of 280 to 450 nm were summed as the total fluorescence intensity at each oil concentration. For light scattering, the summation of measured volume concentrations over droplet sizes of 1.25 to 66.39 μ was used as the total measured oil concentration for each oil sample. Simple linear regression was performed using fluorescence and scattering data of training sets for each oil type, and the resulting linear model was employed to calculate the oil concentration of samples in the test sets. Since having separate calibrations for each oil type is not practical for an online monitoring system, linear models encompassing all four oil types were also evaluated in terms of fluorescence and light scattering.

RESULTS AND DISCUSSION

FLUORESCENCE

The fluorescence intensity of oils depends on oil type. The lube oils 2190 and 9250 had a relatively low intensity, while DFM and JP5 exhibited a relatively high intensity (figure 2). Two of the five excitation wavelengths—266 and 278 nm—produced high fluorescence intensities for all oil types. Only fluorescence spectra based on the 266-nm wavelength were used for all neural network training and testing in this study.

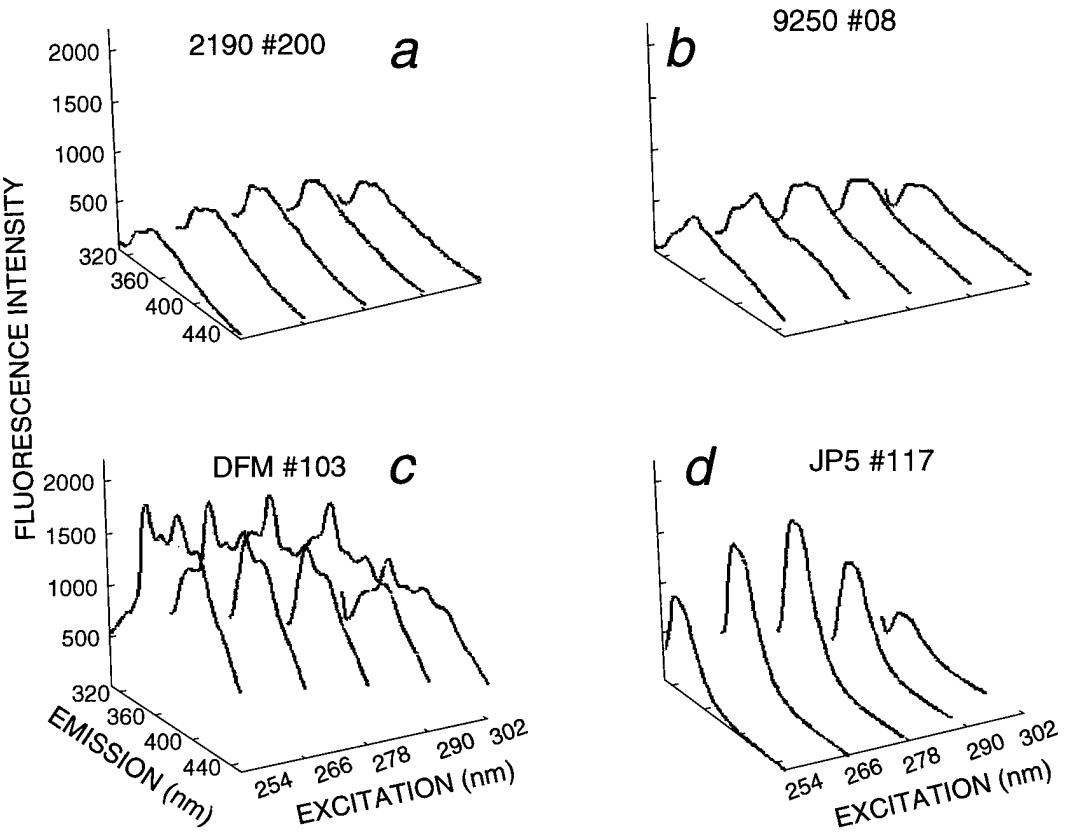

Figure 2. Fluorescence of four oil types at concentration of 50 mg L^{-1} at five excitation wavelengths. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

Figure 3 shows fluorescence spectra of single oils at various concentrations. In general, the spectra were distinguishable at low to medium concentrations. However, the spectral difference between concentrations became smaller when the concentration increased, and the spectra overlapped for JP5 at concentrations greater than 80 mg L^{-1} (figure 3d). Although all four oil types showed different spectral patterns, they were characterized by strong fluorescence in the region of ~350 nm.

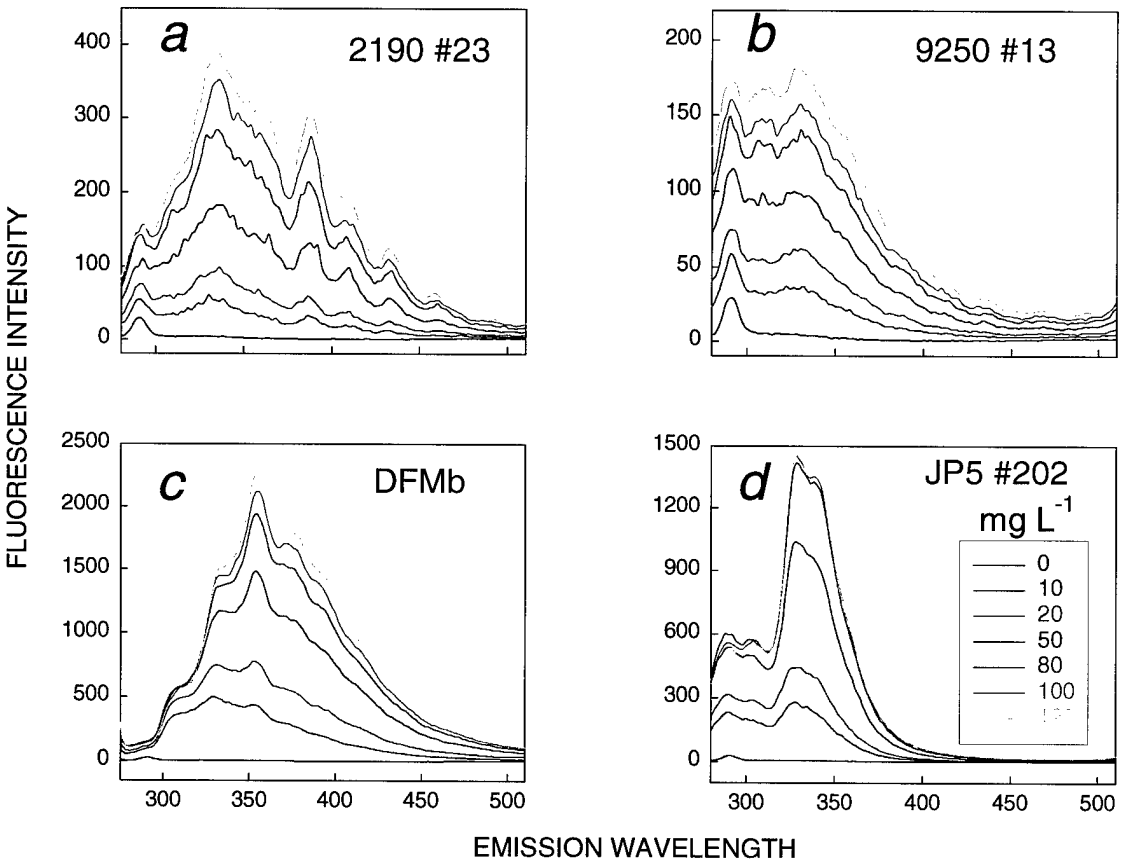

Figure 3. Fluorescence intensity at 266 nm excitation wavelength as a function of oil concentration for four oil types. (a) 2190, (b) 9250, (c) DFM, and (d) JP5. Note that different scales are used for y axes.

Different oil subtypes within the same oil type showed significant differences in fluorescence intensity (figure 4). The greatest variation was found in the lube oil 9250, with a factor of 8 from the lowest to highest (figure 4b). Two DFM oil subtypes, *#208* and *a*, unlike other DFM subtypes, had a strong fluorescence at the wavelength of ~400 nm (figure 4c).

The fluorescence intensity of a mixture typically differs from that obtained by the simple linear addition of its component spectra; however, the difference is relatively small compared to oil types and subtypes (figure 5). The intensity of a mixture may be lower (DFM with 9250, and JP5 with 9250), or higher (9250 with 2190, and JP5 with 2190) than the summation of intensities for the two single oils.

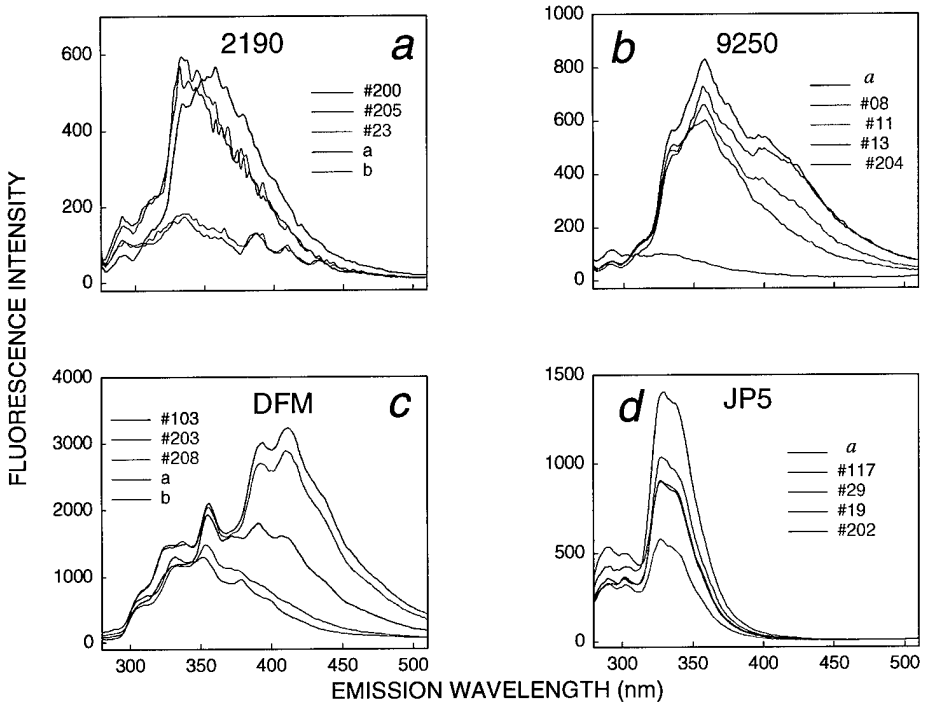

Figure 4. Fluorescence intensity variation with oil subtypes (50 mg L^{-1}) for four oil types. (a) 2190, (b) 9250, (c) DFM, and (d) JP5

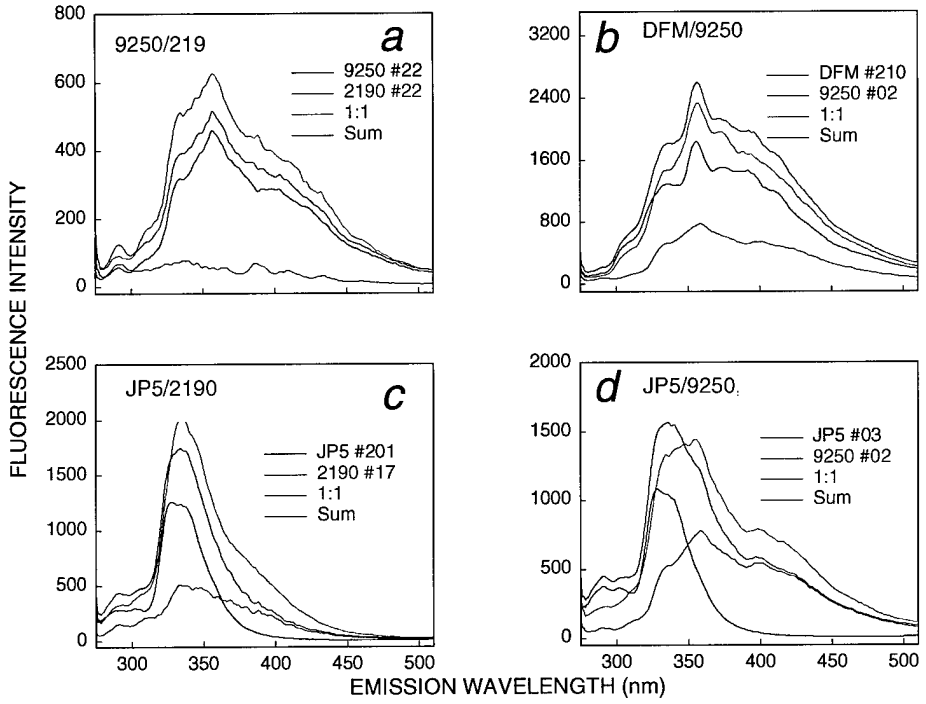

Figure 5. Effect of oil mixing (1:1 ratio) on fluorescence intensity. (a) mixture (60 mg L^{-1}), 9250 and 2190 (both 30 mg L^{-1}); (b) mixture (100 mg L^{-1}), DFM (60 mg L^{-1}), and 9250 (50 mg L^{-1}); (c) mixture (110 mg L^{-1}), JP5 (60 mg L^{-1}), and 2190 (50 mg L^{-1}); and (d) mixture (110 mg L^{-1}), JP5 (50 mg L^{-1}), and 9250 (50 mg L^{-1}).

Wastewater may contain contaminants such as detergents (e.g., Tide® and Mil-D) that influence the fluorescent pattern and intensity of oil samples. Mil-D presented a high fluorescence intensity (maximum ~2300) from 280 to 350 nm, whereas Tide® exhibited a low fluorescence response (maximum ~200) located between 400 and 500 nm (figure 6). The high fluorescence of Mil-D is expected to greatly interfere with the spectral pattern and intensity of oils.

The spectral intensity of DFM*b*, when added with Tide® (60 mg L^{-1}) did not change significantly, although the addition of Tide® slightly increased the fluorescence intensity at the oil concentration of 100 mg L^{-1} (figure 7a). When Tide® was added to the lube oil 2190b, its fluorescence intensity and pattern changed greatly, with a decrease in intensity from 300 to 400 nm and an increase in intensity between 400 and 450 nm at all concentrations examined, which is attributed to Tide® (figure 7b). The detergent Mil-D showed a very different influence on the spectra of DFM*b* and 2190*b*. Mil-D significantly increased the spectral intensity of DFM*b* at ~350 nm at all concentrations (figure 7c). The band at 300 nm resulted from Mil-D itself. The addition of Mil-D to 2190*b* dramatically changed the spectral patterns (figure 7d) so that the high-intensity band of Mil-D at ~300 nm obscured the fluorescence spectra of 2190*b*.

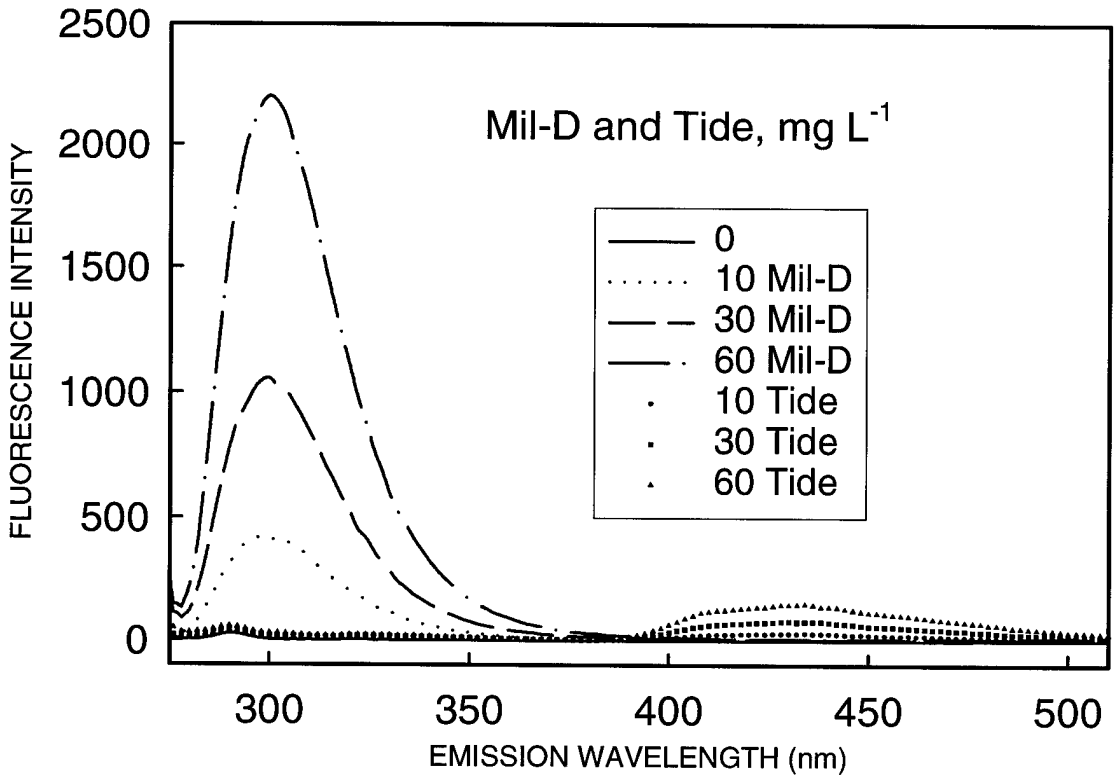

Figure 6. Fluorescence spectra of Mil-D and Tide® at 266 nm excitation wavelength.

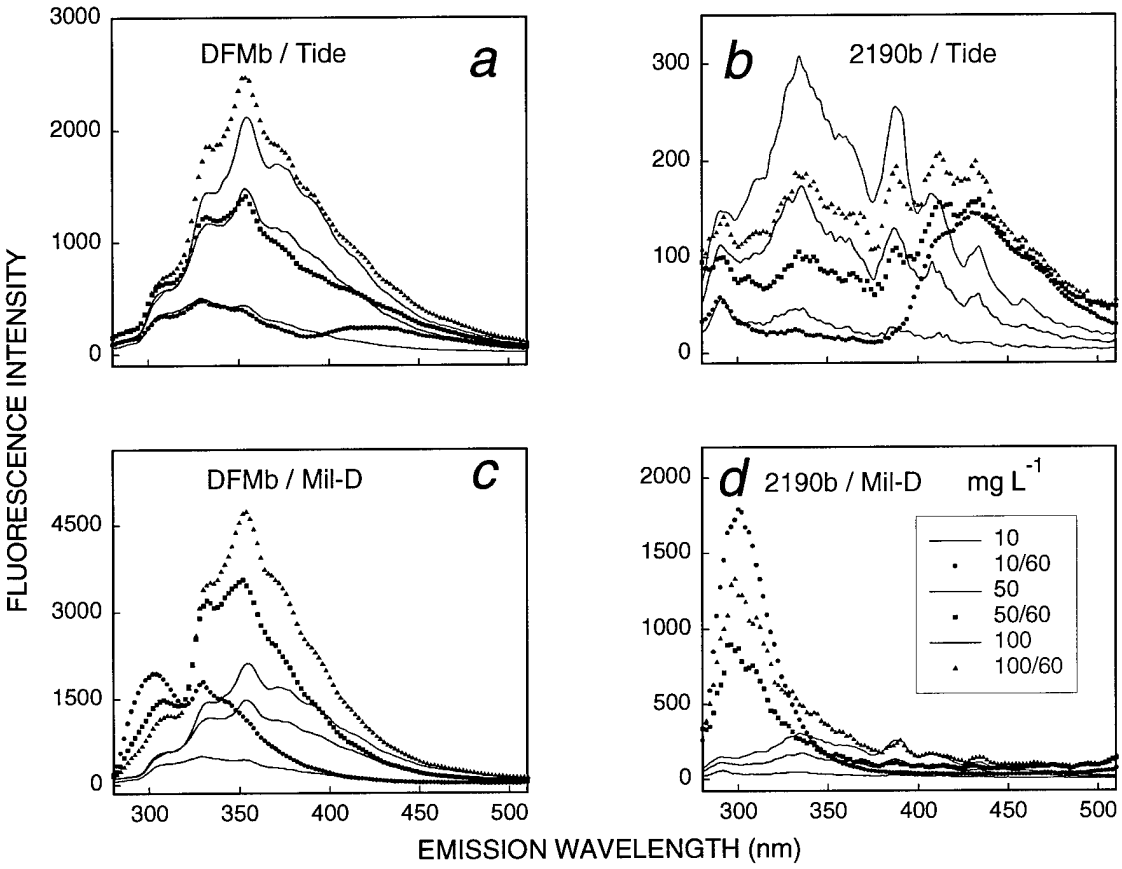

Figure 7. Effect of Mil-D and Tide® on fluorescence spectra of oils. (a) DFM with Tide®, (b) 2190 with Tide®, (c) DFM with Mil-D, and (d) 2190 with Mil-D.

Salt concentration also influences the oil fluorescence. Seawater did not exert a significant impact on the fluorescence of 2190*a*, 9250 *#22*, and DFM*a*, but greatly changed the fluorescent intensity of JP5 (figure 8). When seawater was used instead of deionized water, the intensity of JP5 decreased (figure 8d).

LIGHT SCATTERING

Oil droplet size distribution varied with oil type, and there were generally two size regions, i.e., 1.25 to 3 and 3 to 20 μ except for lube oil 9250, which was predominated by the size of 1.25 to 3 μ (figure 9). The lube oil 2190 and JP5 exhibited distinguishable patterns at all concentrations examined (figures 9a and 9d). For the lube oil 9250, scattering patterns were separable at concentration lower than 50 mg L^{-1}, while a higher concentration resulted in a zero distribution pattern. DFM*b* had a zero distribution pattern at the concentration of 120 mg L^{-1}.

Oil droplet size distribution patterns are similar for oil subtypes with regard to oil types of 9250 and DFM, whereas their intensity (measured volume concentration) varied with subtypes (figure 10). Interestingly, the measured total volume concentration, which was the summation of all sizes from 1.25 to 66.39 μ, was close to the true concentration for the lube oil 2190 and JP5 (figure 11). The concentration of DFM was underestimated. The variation of measured concentrations for oil subtypes

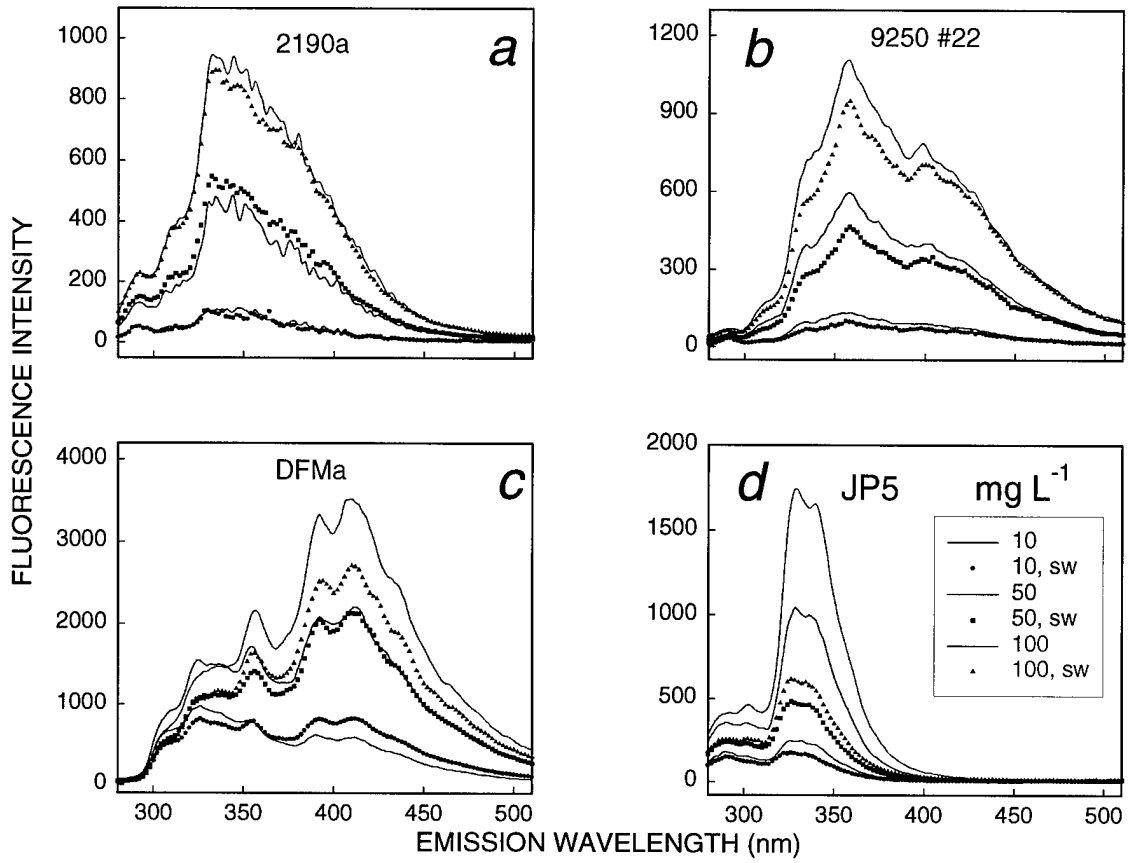

Figure 8. Effect of seawater (100%) on fluorescence spectra of oils. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

were relatively small for the three oil types (2190, DFM, and JP5) (figure 11). However, the variation of measured concentrations for lube oil 9250 was tremendous, with a difference of a factor of 3 among oil subtypes (figure 11). This suggests that it is difficult to calibrate this light scattering instrument using traditional methods.

The volume concentration distribution of an oil mixture was not equal to the summation of the two single oils (figure 12). The measured concentration of the mixture (60 mg L^{-1}) of 9250 and 2190 was significantly higher than the addition of the single oils 9250 and 2190 (30 mg L^{-1} each) (figure 12a). For the other three mixtures (DFM/9250, JP5/2190, and JP5/9250), the volume concentration dropped to zero when 50 mg L^{-1} of single oils was mixed (figures 12b through 12d).

Figure 13 shows the droplet size distributions for two detergents, Mil-D and Tide®. The droplet size for both detergents ranged from 10 to 80 μ, but Mil-D exhibited a much higher volume concentration than Tide®. When Mil-D was added to DFM*b*, the measured volume concentration between 1.25 and 10 μ dramatically decreased (figure 14a). While the addition of Tide® increased the volume concentration of DFM*b* at 50 mg L^{-1} for sizes over 1.25 to 3 μ, Tide® decreased the measured concentration of DFM*b* at 100 mg L^{-1} (figure 14b). For 2190, both Mil-D and Tide® decreased the measured volume concentration (figures 14c and 14d).

Seawater increased the volume concentration of 2190 a at the concentration of 50 and 100 mg L^{-1} at sizes ranging from 1.4 to 20 μ (figure 15a), whereas it decreased the volume concentration of 9250 #22 at the concentrations of 10 and 50 mg L^{-1} (figure 15b). For DFM*a* and JP5, seawater decreased the volume concentration at 50 mg L^{-1}, but it increased the volume concentration at 100 mg L^{-1}, which had a zero volume concentration distribution in deionized water (figures 15c and 15d).

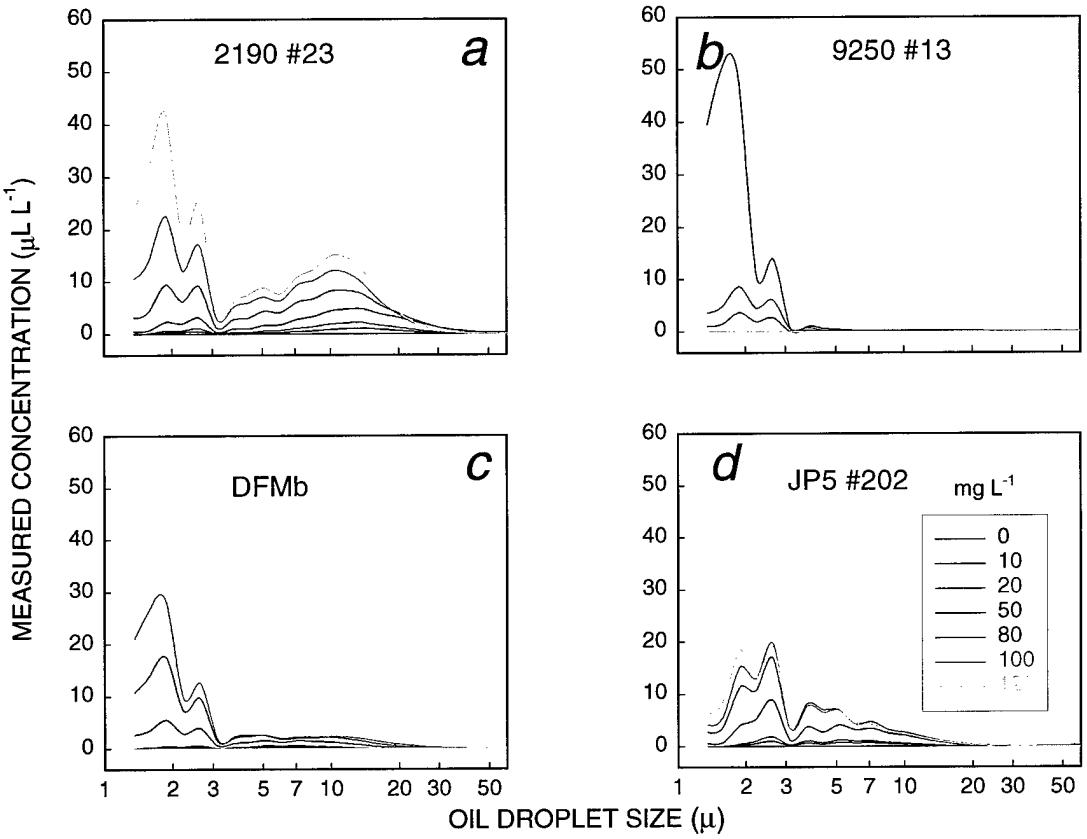

Figure 9. Light scattering as a function of single oil concentration for four oil types. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

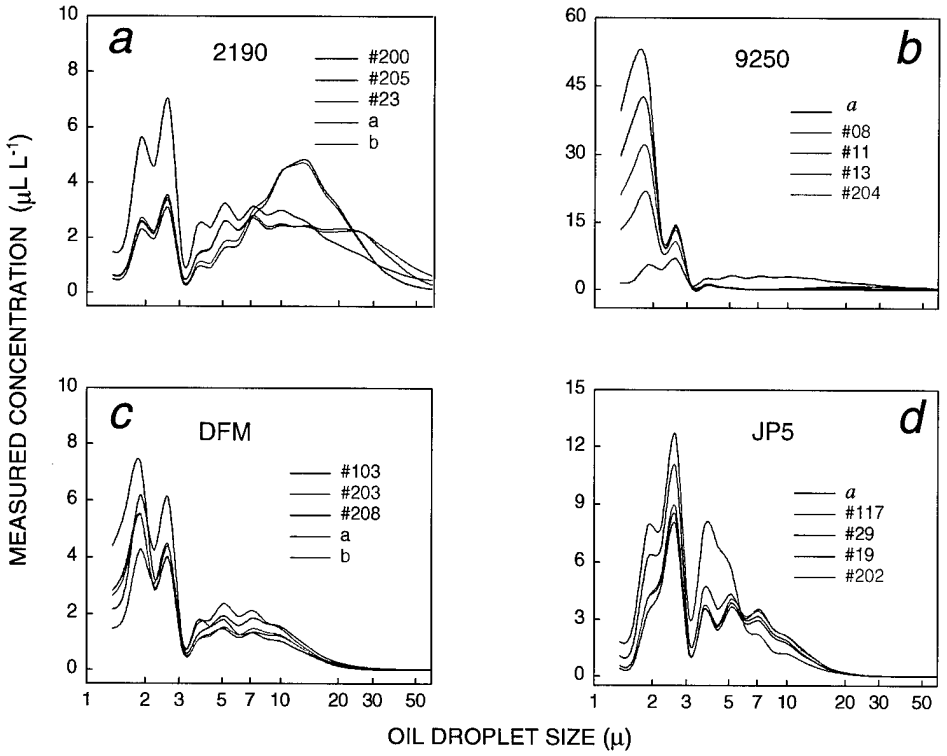

Figure 10. Light scattering variation with oil subtypes (50 mg L^{-1}) for four oil types. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

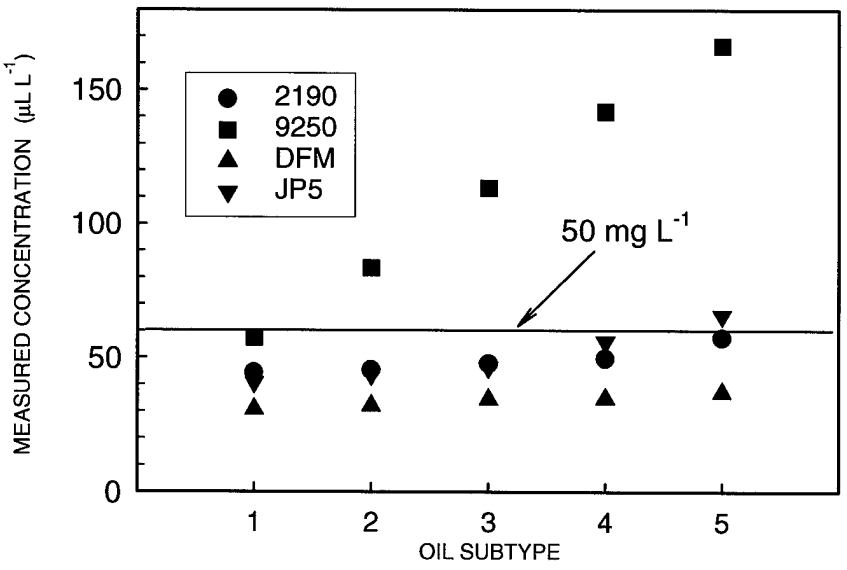

Figure 11. Variation in the total volume concentrations for sizes of 1.25 to 66 μ with oil types and subtypes. Note that the 50 line is used in units of mg L^{-1}, not μL L^{-1}.

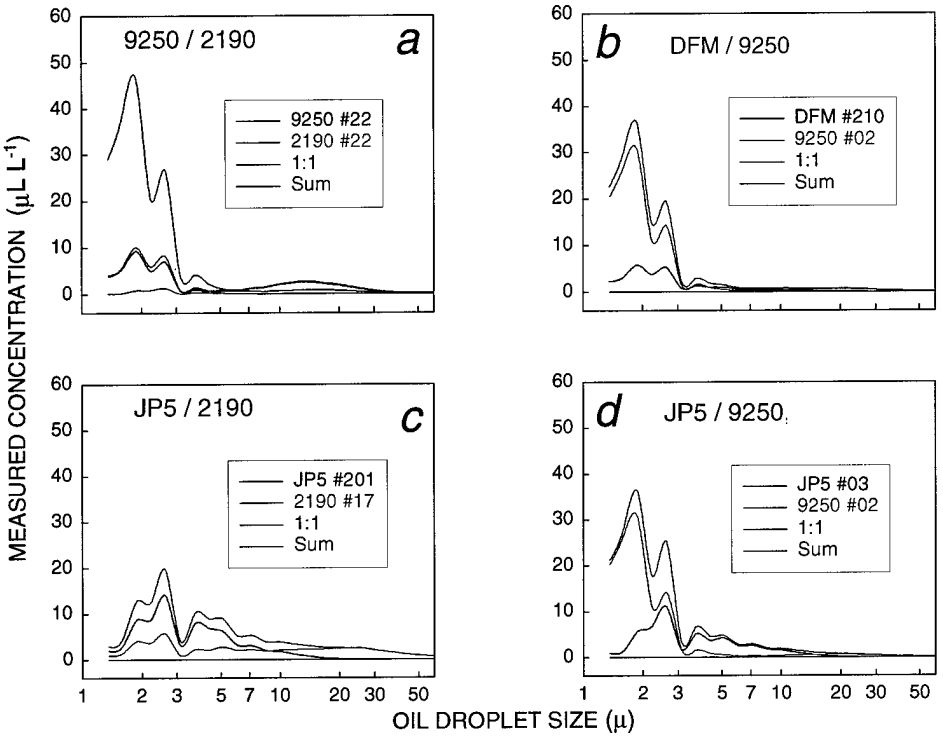

Figure 12. Effect of oil mixing on light scattering. (a) mixture (60 mg L^{-1}), 9250 and 2190 (both 30 mg L^{-1}); (b) mixture (100 mg L^{-1}), DFM (60 mg L^{-1}), and 9250 (50 mg L^{-1}); (c) mixture (110 mg L^{-1}), JP5 (60 mg L^{-1}), and 2190 (50 mg L^{-1}); and (d) mixture (110 mg L^{-1}), JP5 (50 mg L^{-1}), and 9250 (50 mg L^{-1}).

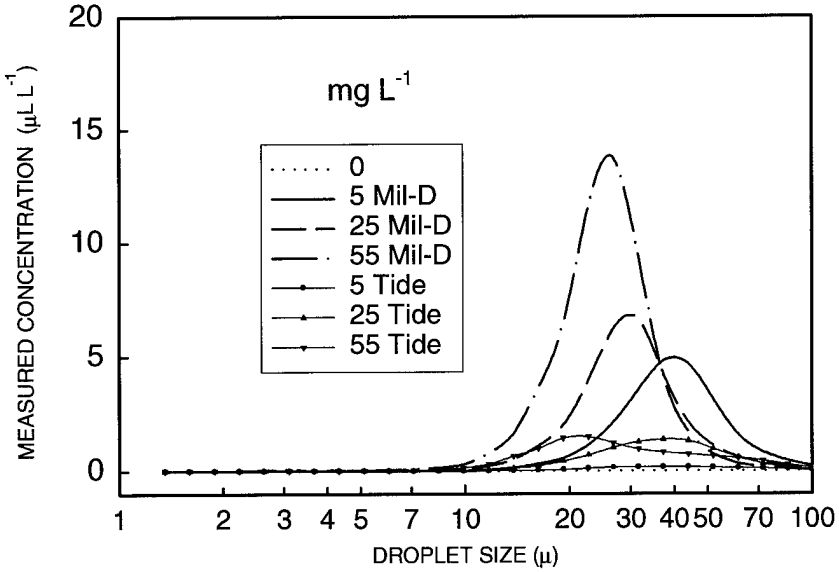

Figure 13. Volume concentration distribution for Mil-D and Tide®.

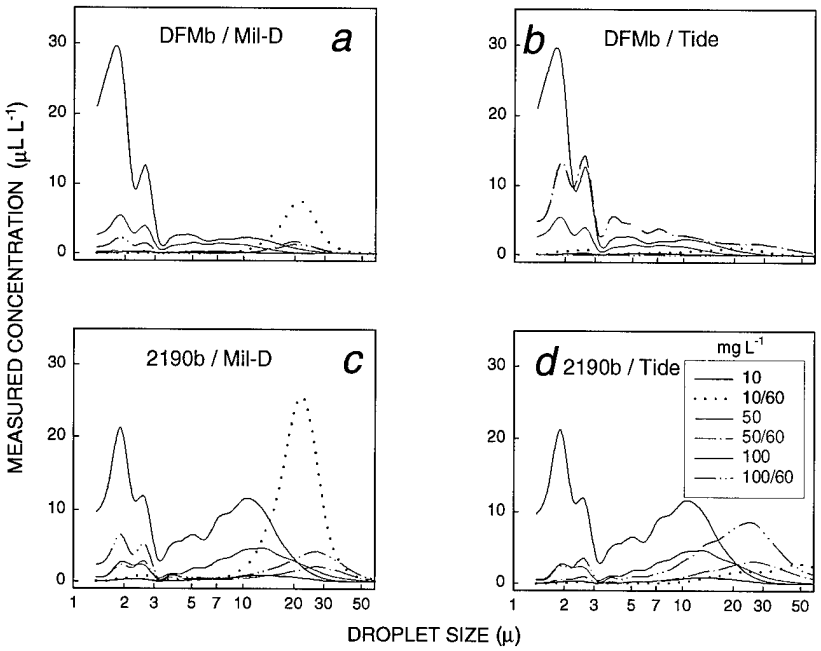

Figure 14. Effect of Mil-D and Tide® on light scattering of oils. (a) DFM with Mil-D, (b) DFM with Tide®, (c) 2190 with Mil-D, and (d) 2190 with Tide®.

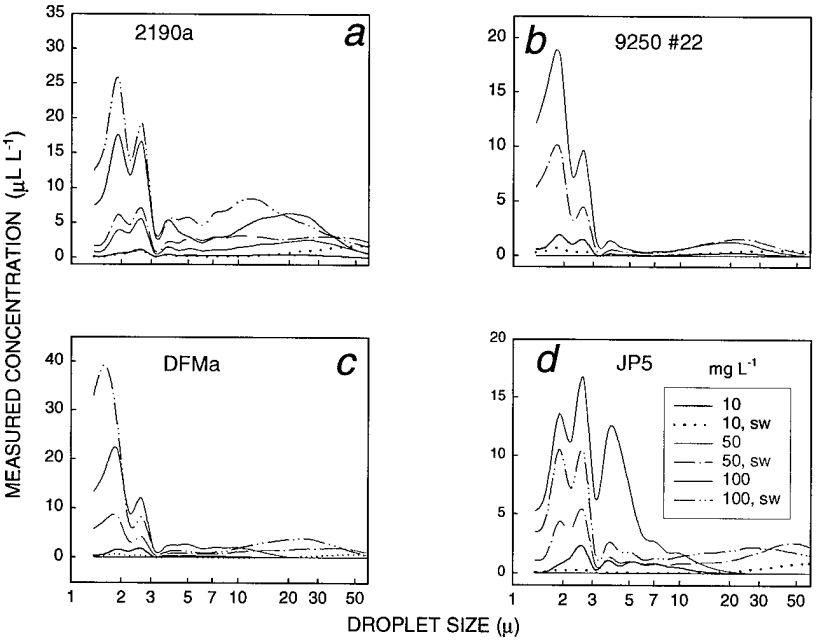

Figure 15. Effect of seawater (100%) on light scattering of oils. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

LINEAR MODEL

For each of the four oil types, there was a reasonable correlation between fluorescence intensity and oil concentration for all single oils, with $r^2 > 0.6$, even though the fluorescence intensity varied very much with the oil subtypes of the same oil type (figure 16). Using the linear models for individual oil types, the prediction of oil concentrations was > 60% except for 9250, which was 40% predicted and mostly overpredicted (figure 17). A poor model was obtained when all of the four oil types was used because of the much greater variation in fluorescence intensity among oil types (figure 18a). With this unified model, only 28% of the oil samples was predicted (figure 18b; table 1).

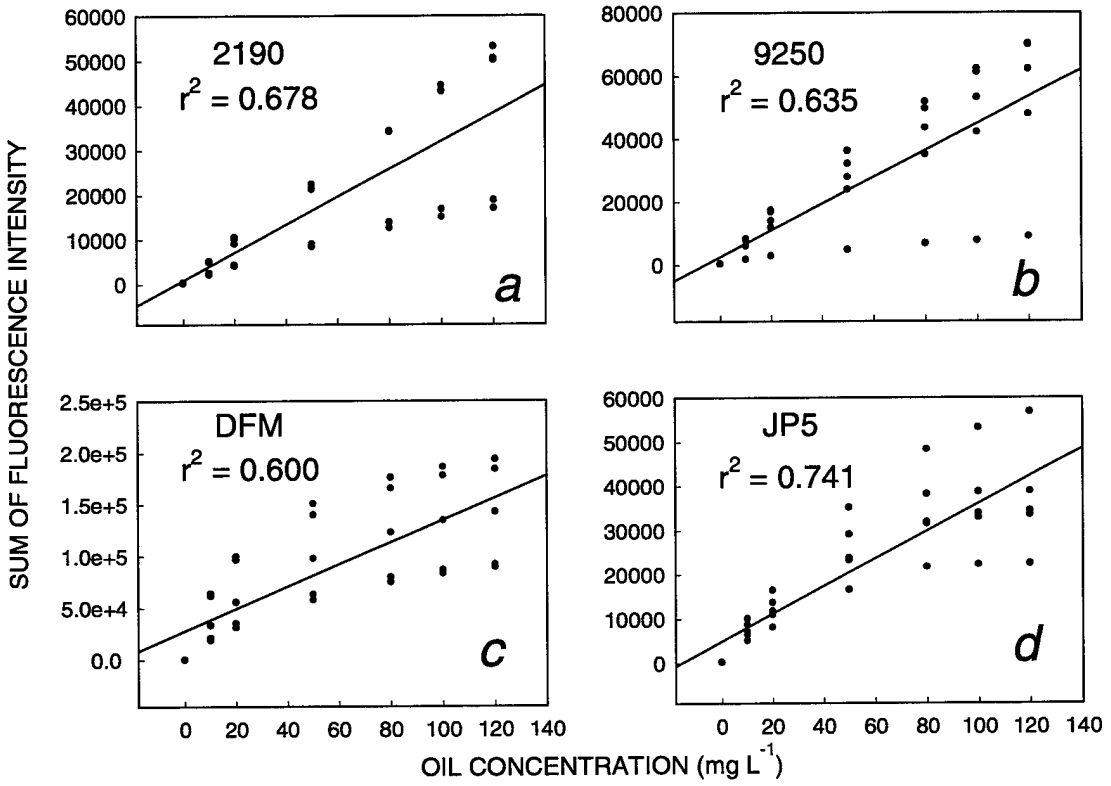

Figure 16. Oil fluorescence intensity as a function of concentration over emission wavelengths of 280 to 450 nm. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

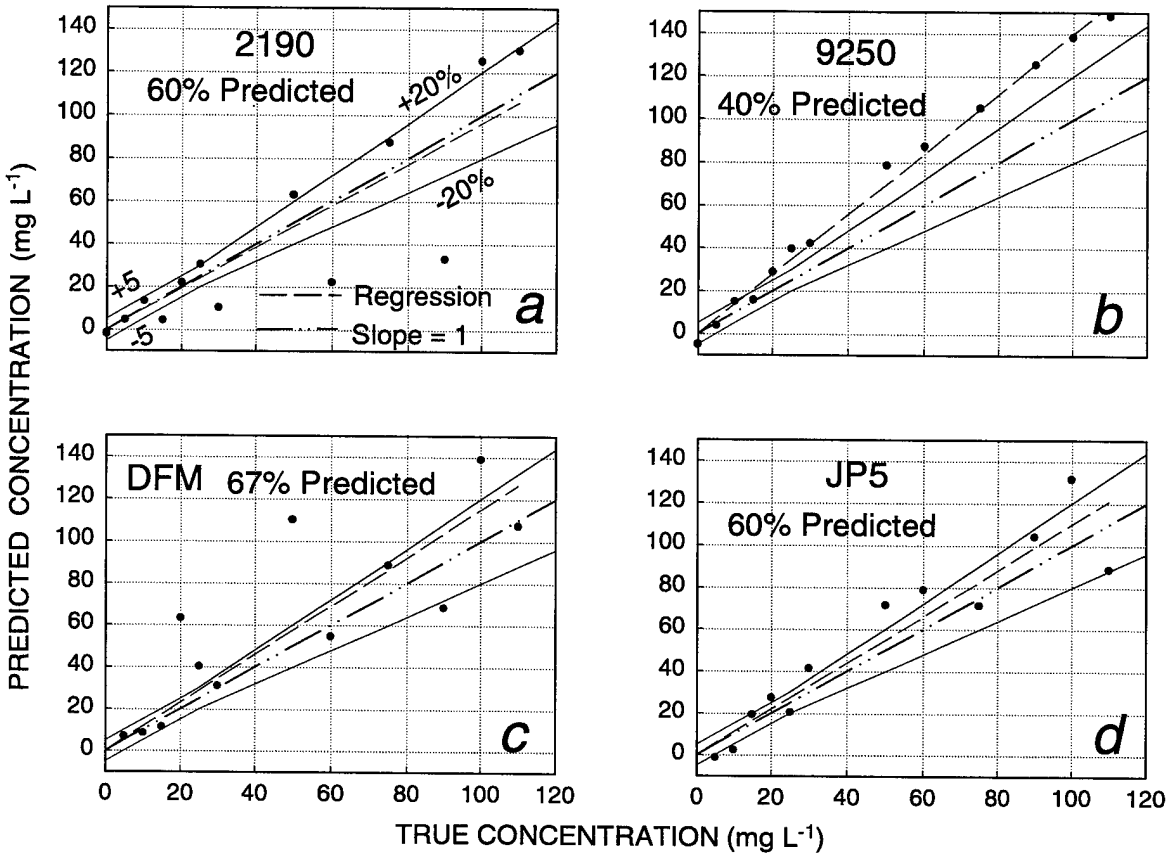

Figure 17. Oil concentration prediction using the linear calibration of fluorescence for individual oil type. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

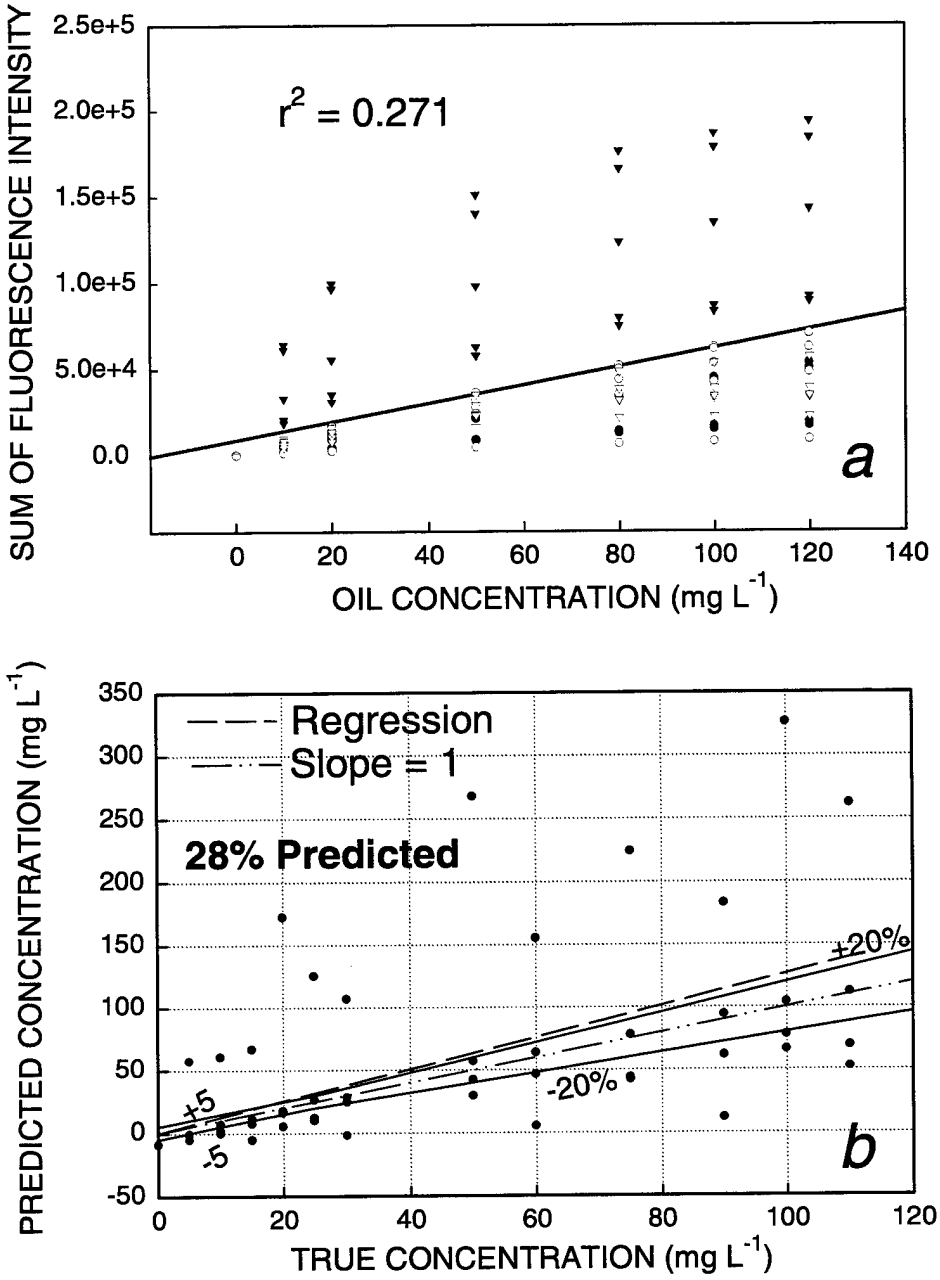

Figure 18. Oil concentration prediction using the linear model of fluorescence intensity for all four oil types. (a) linear calibration and (b) prediction.

Table 1. Comparison of prediction percentage for single oils between the linear model and the artificial neural network (ANN).

Variable	Fluorescence	Scattering	Fluorescence + Scattering
Linear model	28	20	23
ANN	62	77	100

For light scattering, there was an excellent correlation between the measured volume concentration and the true oil concentration for 2190 (figure 19a), but no correlation was found for 9250 and DFM (figure 19b and 19c), both of which had a zero volume concentration distribution at a high oil concentration. The prediction using the linear models for individual oil types was 60% for 2190, 20% for 9250, 53% for DFM, and 67% for JP5 (figure 20). Using all the scattering data for the four oil types, the correlation between the measured volume concentration and the true oil concentration decreased to $r^2 = 0.281$ (figure 21a), and the prediction was only 20% (figure 21b; table 1).

The use of a multivariate linear model, in which both fluorescence and light scattering data were employed, failed to improve the prediction of the concentration of single oils. The prediction rate was only 23% (table 1).

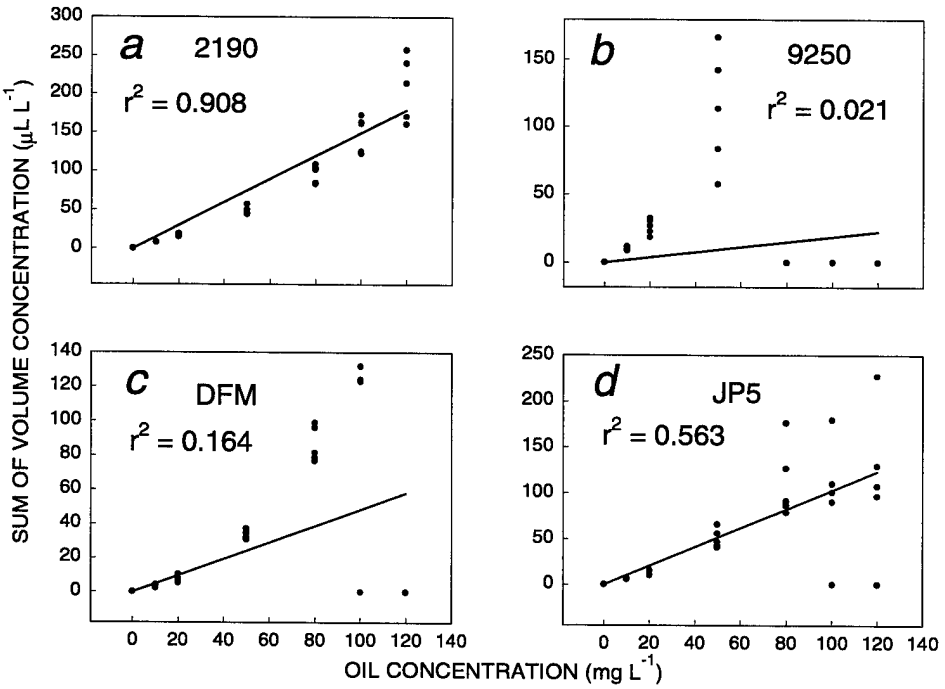

Figure 19. Measured volume concentration as a function of oil concentration over droplet sizes of 1.25 to 66.39 μ. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

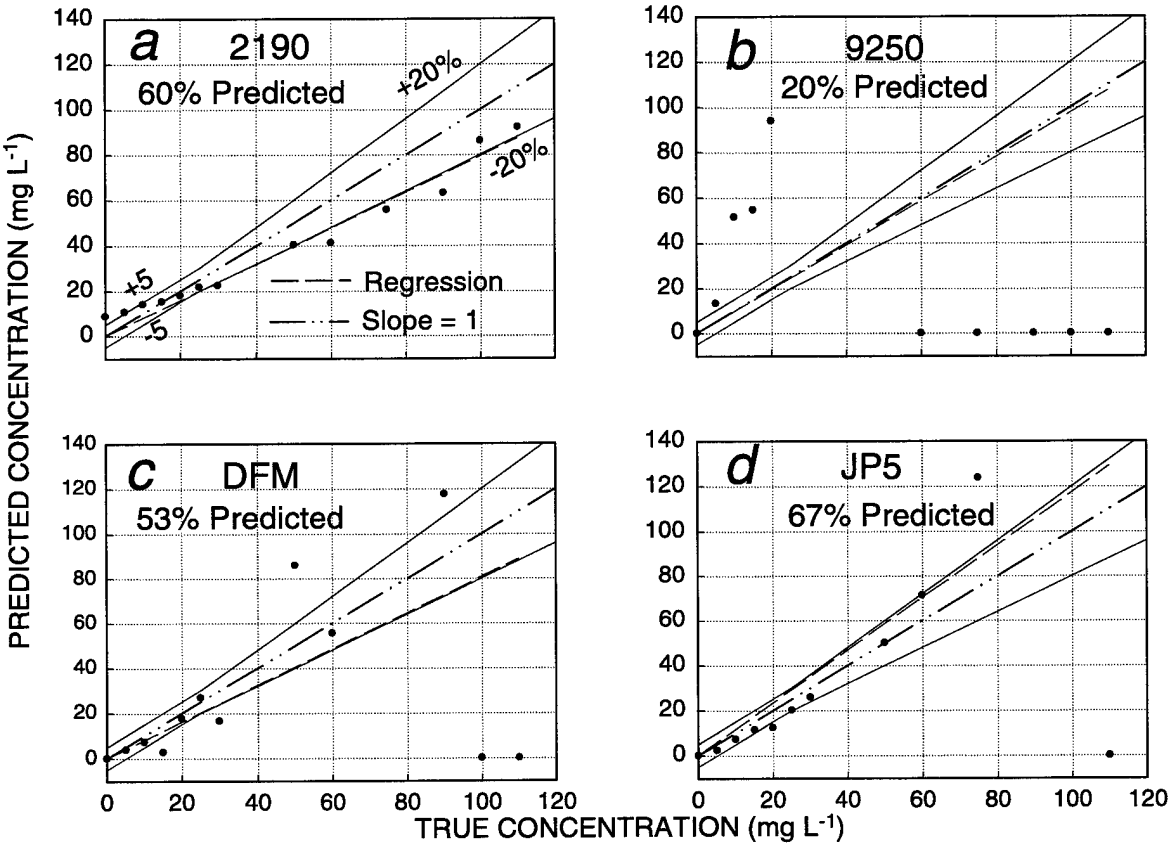

Figure 20. Oil concentration prediction using the linear calibration of light scattering for individual oil type. (a) 2190, (b) 9250, (c) DFM, and (d) JP5.

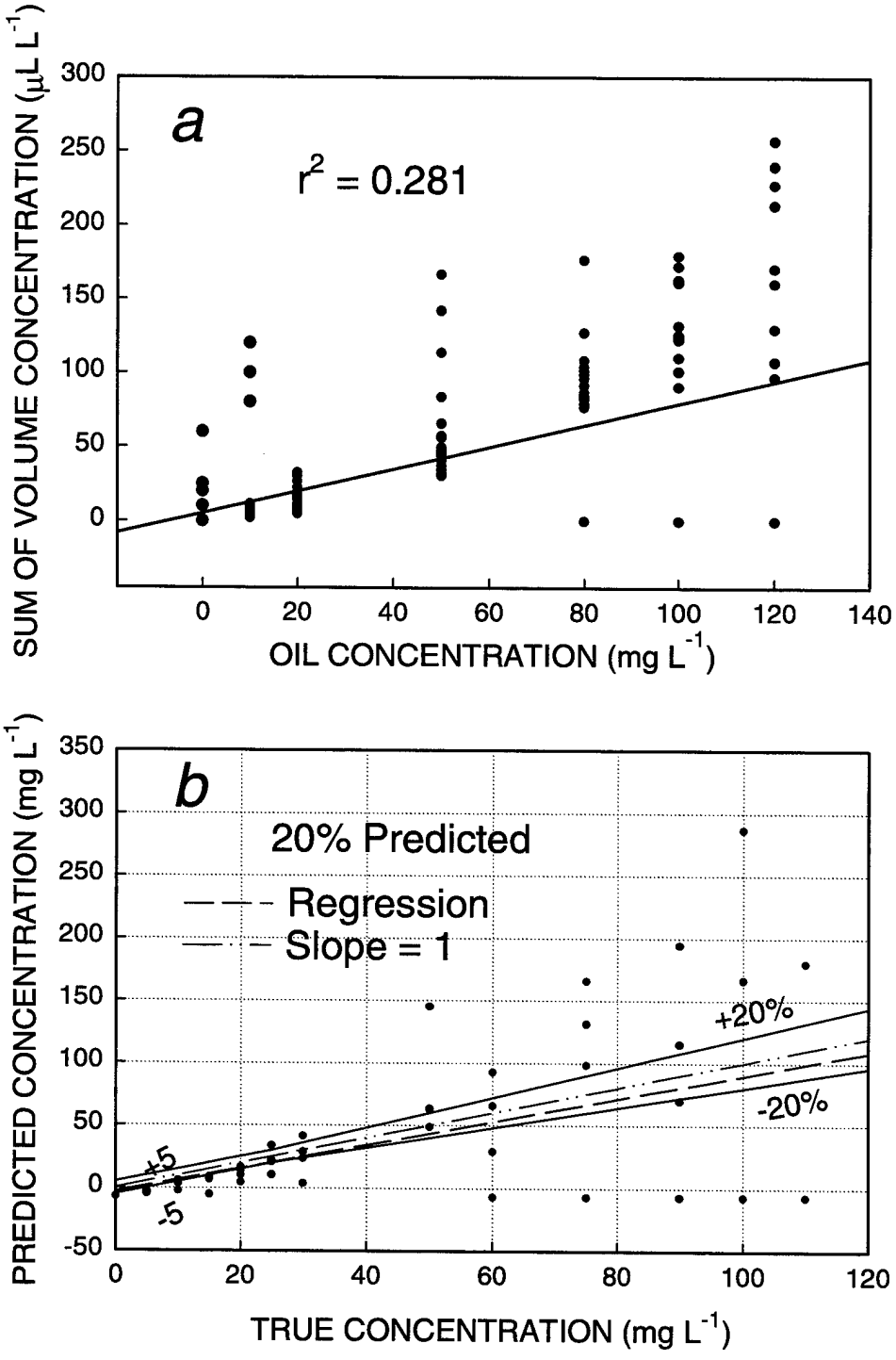

Figure 21. Oil concentration prediction using the linear model of light scattering for all four oil types. (a) linear calibration and (b) prediction.

NEURAL NETWORK OPTIMIZATION

Data sets used to train a network varied with application. The training data set may be one or a combination of seven sample sets. These consisted of single oils (140 samples), mixed oils (84 samples), single oils (2190 and 9250) with Mil-D or Tide® added (213 samples each set), single oils (DFM and JP5) with Mil-D or Tide® added (213 samples each set), and single oils with seawater added (224 samples). There were nine data sets for testing the trained neural network (table 2).

After data sets for training and testing were established, parameters in a neural network were optimized to generate the best architecture of the network. The parameters that were optimized included the number of iterations, the number of neurons in the hidden layer, the number of inputs and outputs, learning rate, and momentum term. The data sets of single oils were used as an example to show how these parameters were optimized.

Table 2. Prediction percentage relevant to training and test data sets.

Train Set	Test Set								
	Single (60)#	Mixture (84)	Single /Mil-D (2190 & 9250) (89)	Single /Tide® (2190 &9250) (90)	Single /Mil-D (DFM & JP5) (90)	Single /Tide® (DFM & JP5) (90)	Single /sw (60)	Mix /Mil-D (84)	Mix /Tide® (84)
Single oils (140)#	100	92	40	36	40	69	59 (224)		
Mixture (84)	n.p.	90	n.p.	n.p.	n.p.	n.p.	n.p.		
Single/Mil-D (213) (2190 & 9250)	52	46	79	46	44	42	45 (224)		
Single/Tide® (213) (2190 & 9250)	48	41	24	93	27	41	63 (224)		
d (852)	78	73	78	93	86	96	82 (224)	74	68
S+d (992)	97	93	74	90	86	92	91	73	69
S+sw+d (1216)	92	75	72	90	84	93	100	73	60
S+d+m (1076)	97	96	71	92	88	98	91	74	73
S+sw+d+m (1300)	93	88	78	90	87	92	100	74	67

# Number of training or test pairs in the data set.
S = single oil, d = detergent added, sw = seawater added, m = mixed oil.
n.p. Not predicted.

Number of Inputs and Outputs

The number of inputs depends on data available for the network. The input data in this study can be all the values of a full emission spectrum of fluorescence and light scattering, intervalically selected values, or characteristic bands of a spectrum. The whole spectrum consists of 151 values for fluorescence and 34 values (including two transmittances) for light scattering. The selection of characteristic bands may be the ideal choice for a single oil or a mixture of known components, but it is not applicable in practice for mixtures, the components of which are usually unknown. The whole spectrum or regularly selected values are therefore practical candidates as input data. Since using a whole spectrum was not superior to selected values in terms of concentration prediction, the intervalically selected spectral values (every 10 nm) of fluorescence were used as input data in this study. The number of values was 31. For scattering, only the first 24 values were selected, which included oil droplet sizes up to 66 μ. Droplets greater than 66 μ in a sample are considered as air bubbles. In addition, two transmittances from scattering were also included as input data. One is the laser transmission that is the measured laser power, in mW, transmitted through a sample, and the other is the optical transmission related to the ratio of the transmitted laser power to the original laser power.

Three different groups of training data sets, i.e., fluorescence data (31 inputs), scattering data (26 inputs), and combination of the two (57 inputs), were used to determine which group was the best (table 1). The prediction based on fluorescence was the poorest (62% predicted) (figure 22a), with scattering second (77% predicted) (figure 22b), and a combination of the two the best (100% predicted) (figure 22c). For the network trained with the fluorescence data, concentrations were mostly under-predicted (slope = 0.754). Therefore, all the networks presented in this report were trained using the combination of fluorescence and light scattering data.

The network had one output (i.e., the predicted concentration of an oil sample).

Number of Iterations

The backpropagation network may be over-trained, i.e., the network gives perfect results for the training set (low RMS), but poor predictions for the test set (high RMS) (Zupan and Gasteiger, 1991). While training iterations increased from 1,000 to 40,000, RMSs for both the training and test decreased, and the prediction rate increased (figure 23). When the network was trained more than 80,000 iterations, the training RMS kept decreasing, but the test RMS remained the same, and the prediction rate decreased. The optimized number of training iterations for this specific example was 80,000, at which the prediction rate was 100% (figure 23).

Number of Neurons in Hidden Layer

A number was chosen after several trials, which demonstrated that the network performed no better with additional neurons. When choosing the number of neurons, it is important to make sure there are enough neurons to allow the network to learn its task. If the network performs significantly better with only a few additional neurons, this is an indication that too few hidden neurons are being used initially. However, if the network performance does not change with additional hidden neurons, then the initial amount is a safe number to use. The number of neurons in the hidden layer influenced the prediction performance (figure 24). The prediction rate for the testing data set increased with an increase in the number of hidden neurons. When the number of hidden neurons exceeded 20, the prediction percentage decreased. It is evident that the network's prediction performance was the best when 20 hidden neurons were used (figure 24). Too few neurons may give the network too few

adjustable weights so that the nonlinear properties of the system are not fully used. Too many neurons might make the network too complex to model the actual data set by introducing additional noise into the network.

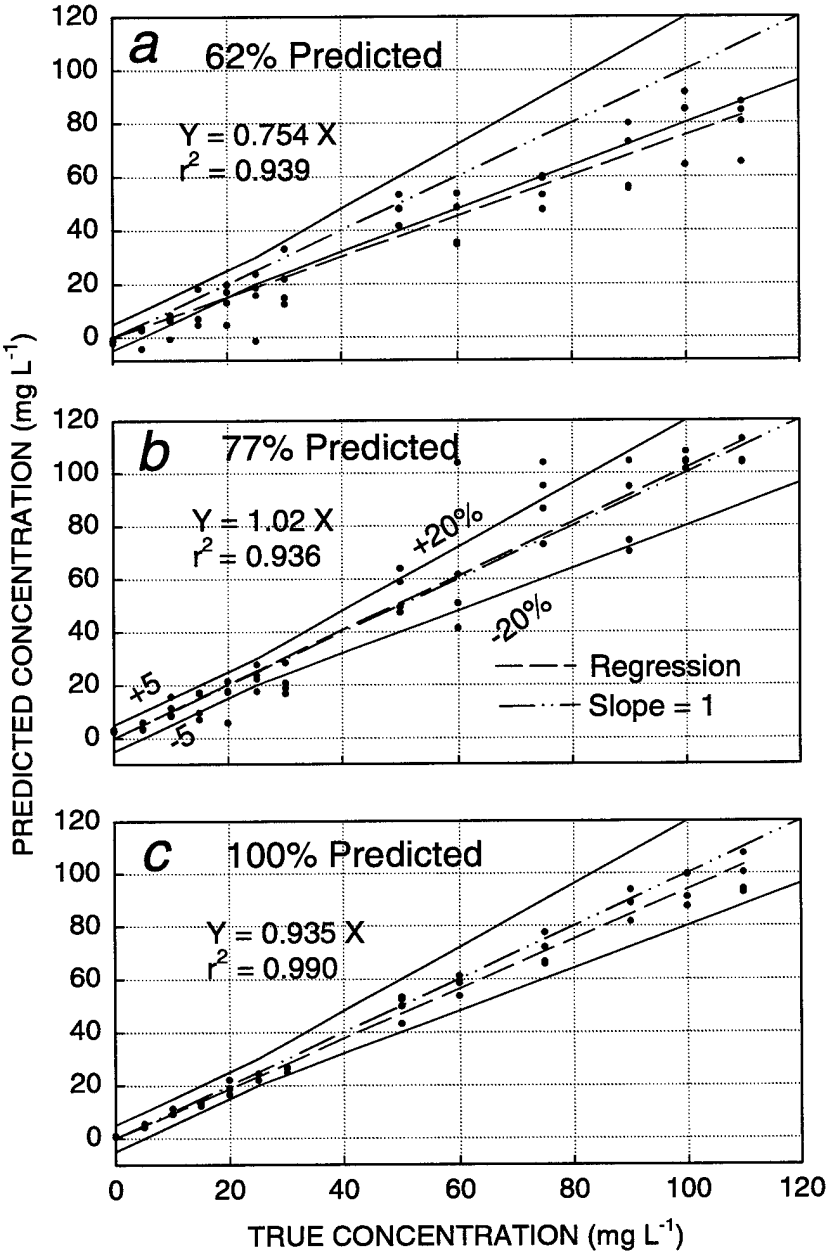

Figure 22. Prediction of single oil concentrations using three networks trained with different data sets. (a) fluorescence data only, (b) scattering data only, and (c) fluorescence plus scattering.

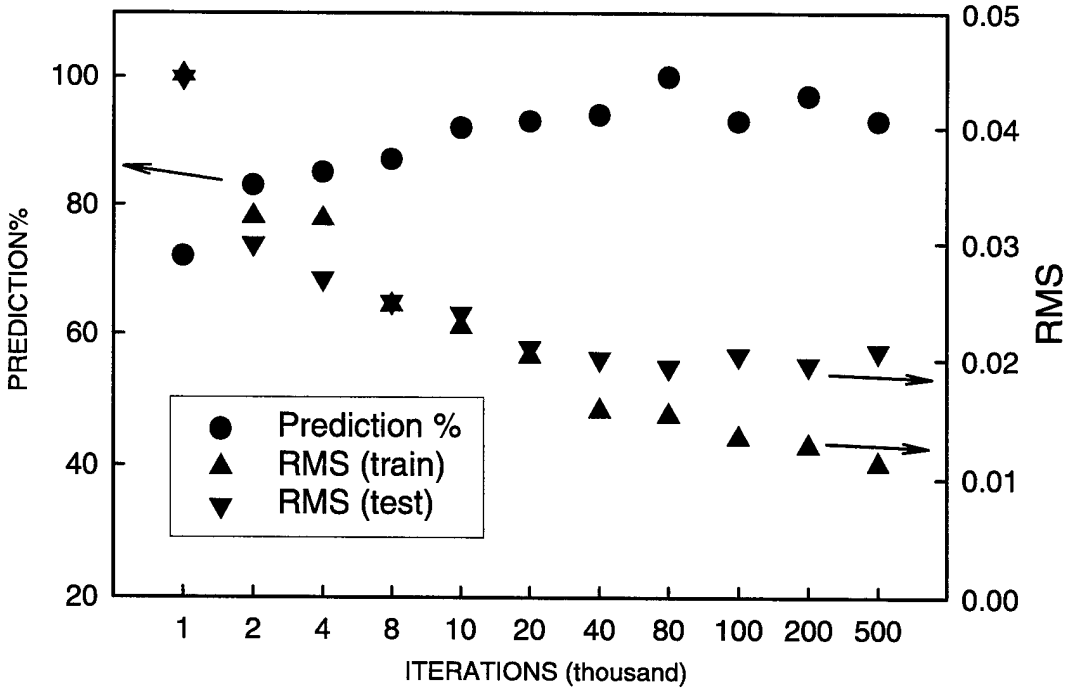

Figure 23. Prediction percentage and RMS as a function of training iterations.

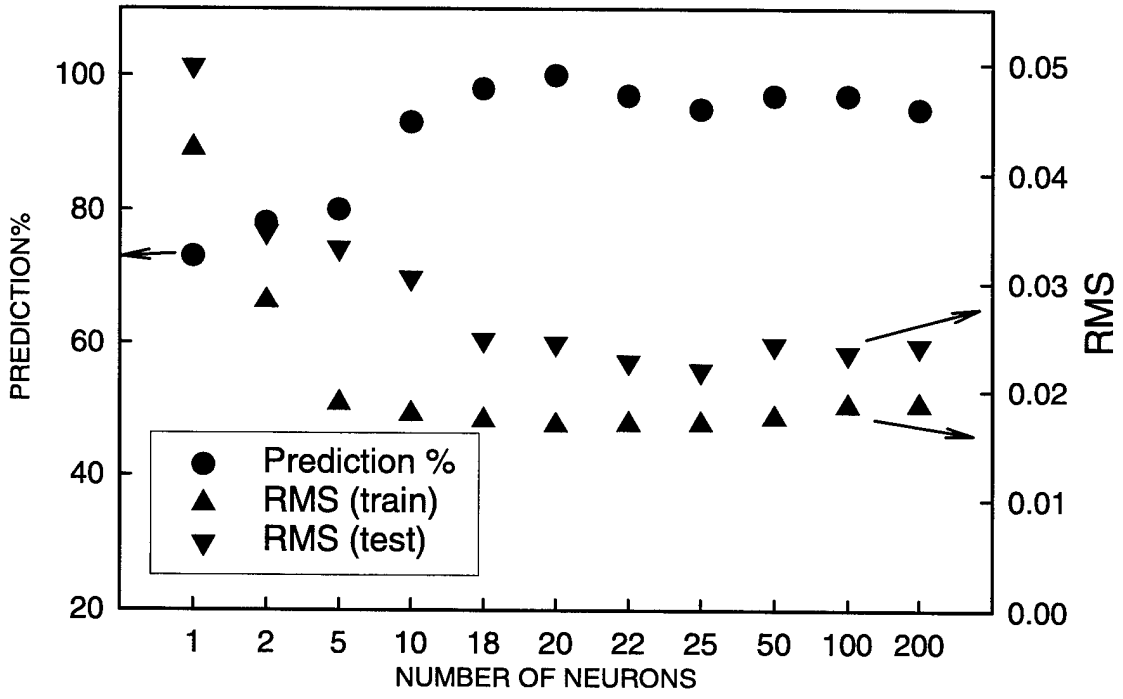

Figure 24. Prediction percentage and RMS as a function of number of neurons in hidden layer.

Learning Rate

Learning rate (η) is another important parameter that can be adjusted during the training process. A learning rate that is too large leads to unstable learning, whereas a learning rate that is too small results in excessively long training time. In general, a larger rate is used at the beginning of the network's training, and a smaller value is used when the network approaches the convergence point.

Changing both learning rates for the hidden and output layer affected prediction performance as well as RMS errors (figures 25 and 26). While the learning rate for the hidden layer increased from 0.1 to 20, RMS (train) monotonically decreased, whereas the RMS (test) decreased at learning rates from 0.1 to 5, but increased beyond 5. The learning rate of 5 produced a 100% prediction. Similar results were found for the output layer. RMSs for both training and test increased and prediction rate decreased when the learning rate for the output layer exceeded 0.1 (figure 26). The combination of 5 and 0.05 for the hidden and output layer, respectively, generated the best prediction.

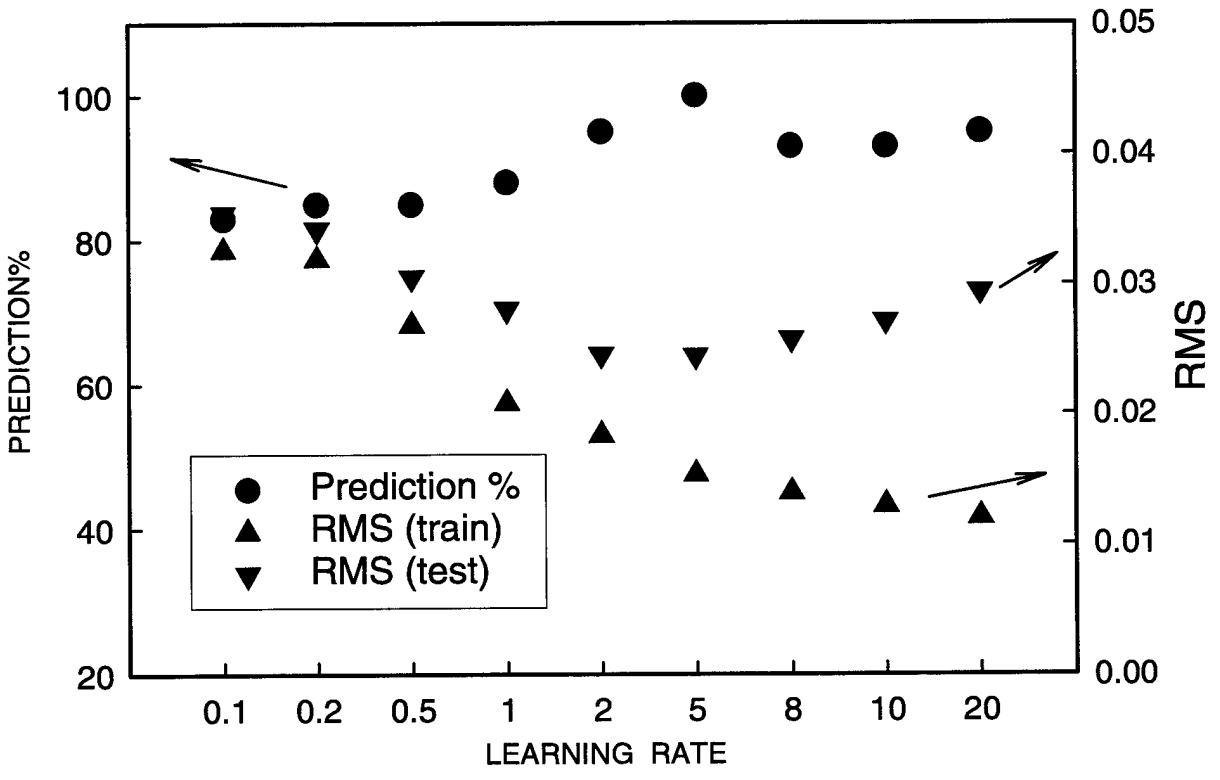

Figure 25. Prediction percentage and RMS as a function of learning rate for hidden layer.

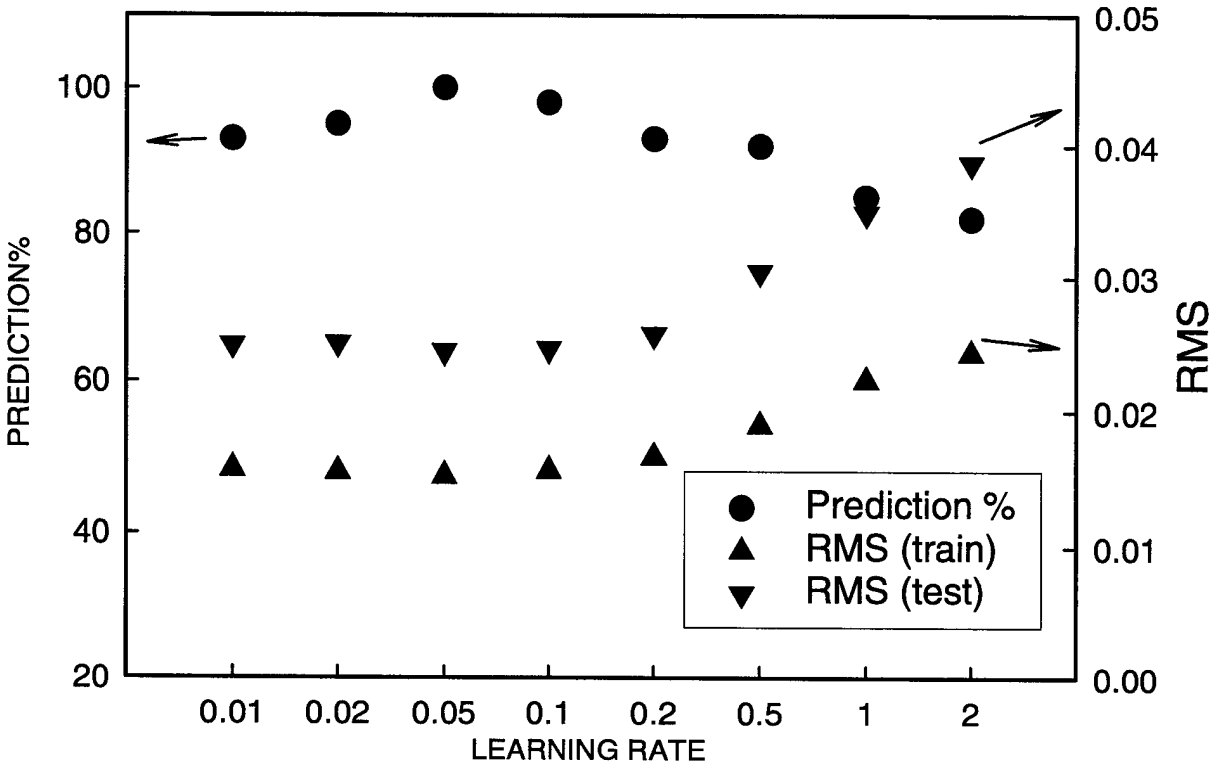

Figure 26. Prediction percentage and RMS as a function of learning rate for output layer.

PREDICTION OF OIL CONCENTRATIONS USING VARIOUS TRAINED NEURAL NETWORKS

For this study, the neural network could be trained with any combination of the seven individual training data sets. Literally speaking, there were 127 combinations of training data sets. Each network trained with one combination can be used to test the nine test data sets (table 2). In general, the trained network gives good predictions for test data sets similar to the training data sets. Whenever a network was trained in the presence of single oils, single oil concentrations were predicted >90%. The concentrations of 84 mixed oil samples were predicted well (92% predicted) using 140 single oil samples as a training data set, with a correlation coefficient of 0.951 (figure 27a). The prediction of mixed oils using mixed oils as a training data set was 90% (figure 27b). However, when the combination of single and mixed oil samples was used to train the network, the mixed oil prediction was improved to 98% (figure 27c).

Based on prediction percentage, the best network trained to predict all the nine test data sets was the combination of single oils, mixed oils, and single oils with detergent added (table 2). This trained network predicted oil concentrations very successfully (88 to 98% predicted) except for the three sets of test data: 2190 and 9250 with Mil-D added, mixed oils with Mil-D added, and mixed oils with Tide® added. The worst group was lube oil 2190 and 9250 with Mil-D added, with the prediction percentage of 71% (figure 28). This is caused by the extremely high impact of Mil-D on the signals of fluorescence and scattering of 2190 and 9250, both of which possess a relatively low intensity of fluorescence. It should be pointed out that most of the incorrectly predicted were samples with low oil concentrations (5 or 10 mg L^{-1}) or high Mil-D concentrations (30 or 60 mg L^{-1}) (figure 28). Since

most of the incorrectly predicted samples had concentrations of 5, 30, 60, and 75 mg L^{-1}, which were not included in the training data set, a well-defined training data set should improve its prediction percentage. All samples (except one) with oil concentrations >90 mg L^{-1} were correctly predicted.

Mixed oils with Mil-D or Tide® added were not predicted as well as other test data sets, i.e., 74% and 73%, respectively (figures 29 and 30). This may have resulted from no similar data sets included in the training data set. Note that most of the incorrectly predicted samples were samples with high Mil-D or Tide® added (30 or 60 mg L^{-1}).

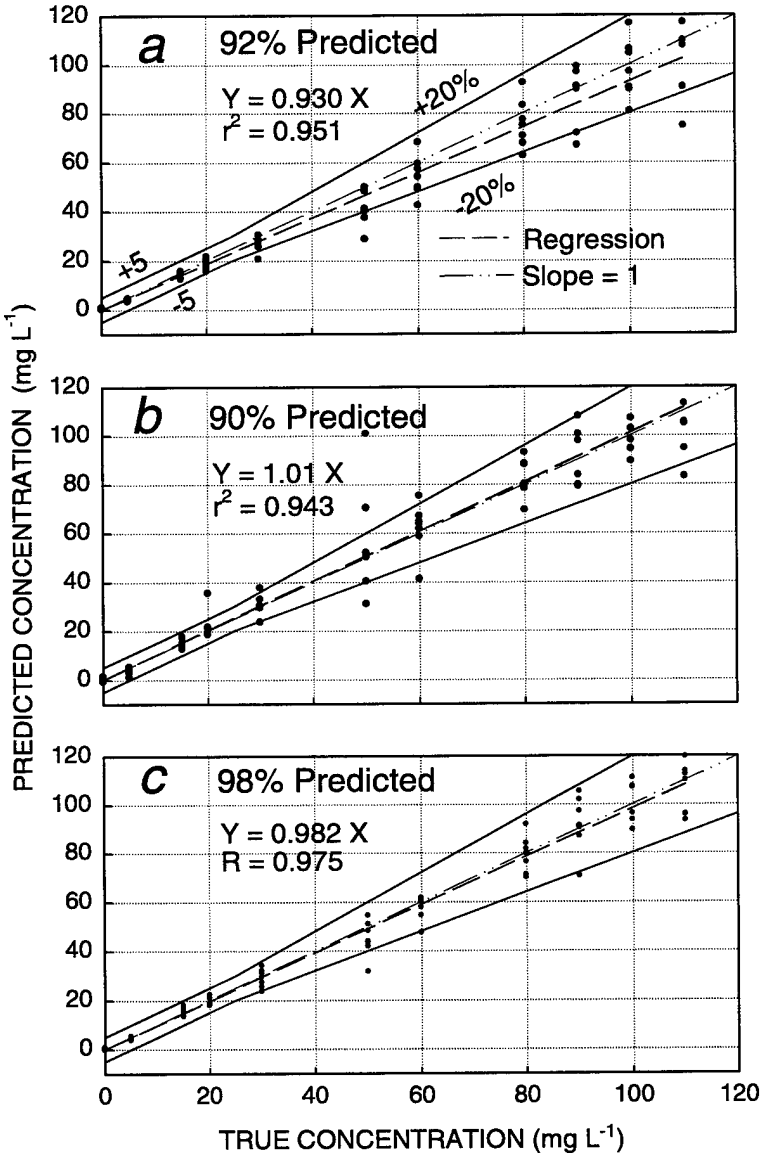

Figure 27. Prediction of mixed oil concentrations using three networks trained with different data sets. (a) trained with single oils, (b) trained with mixed oils, and (c) trained with single and mixed oils.

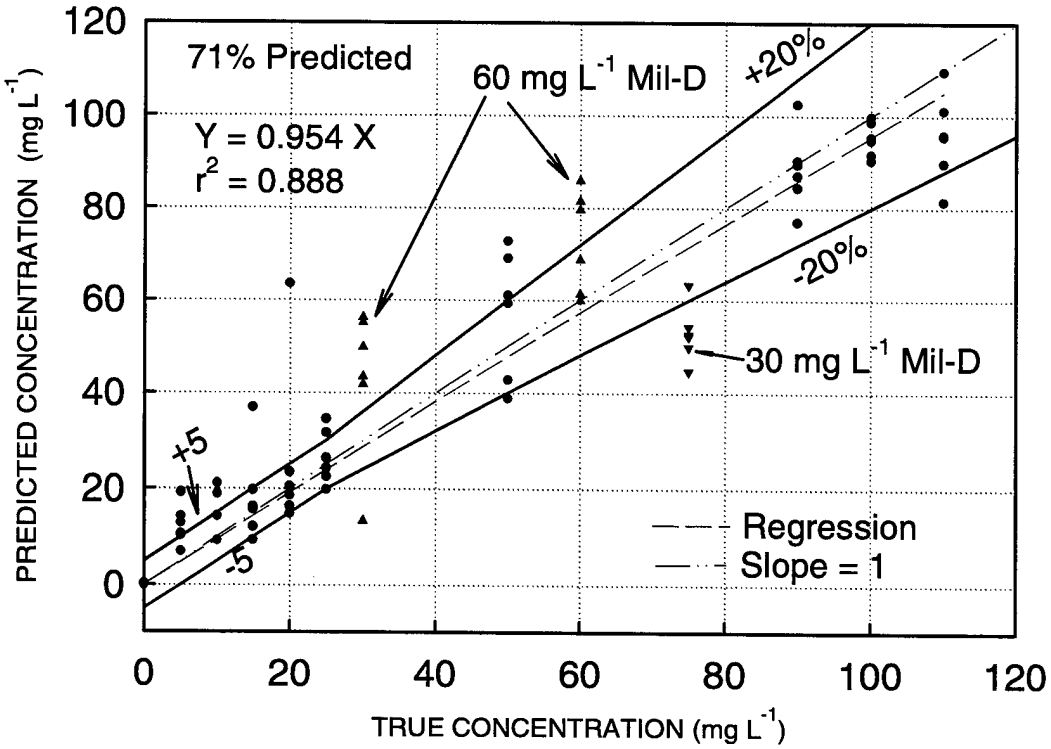

Figure 28. Prediction of oil concentrations for single oils (2190 and 9250) with Mil-D added. Network was trained with a combination of data sets: single oils, mixed oils, and single oils added with detergents. ▲ with 60 mg L^{-1} Mil-D; ▼ with 30 mg L^{-1} Mil-D.

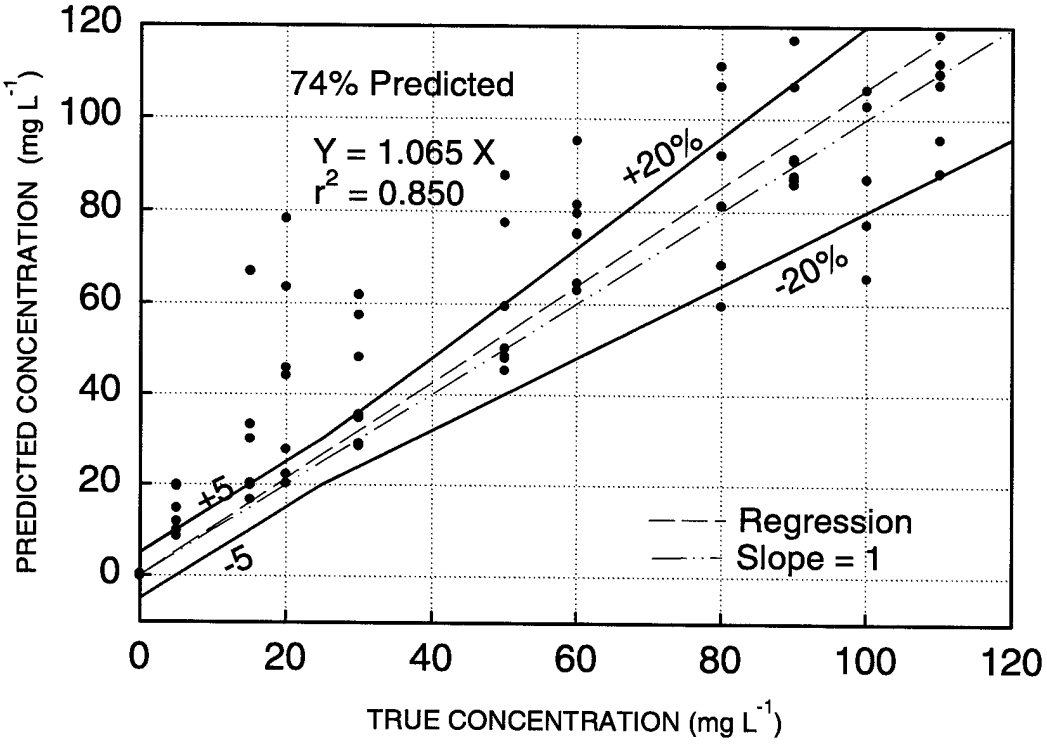

Figure 29. Prediction of oil concentrations for mixed oils with Mil-D added. The network was trained with a combination of data sets: single oils, mixed oils, and single oils added with detergents.

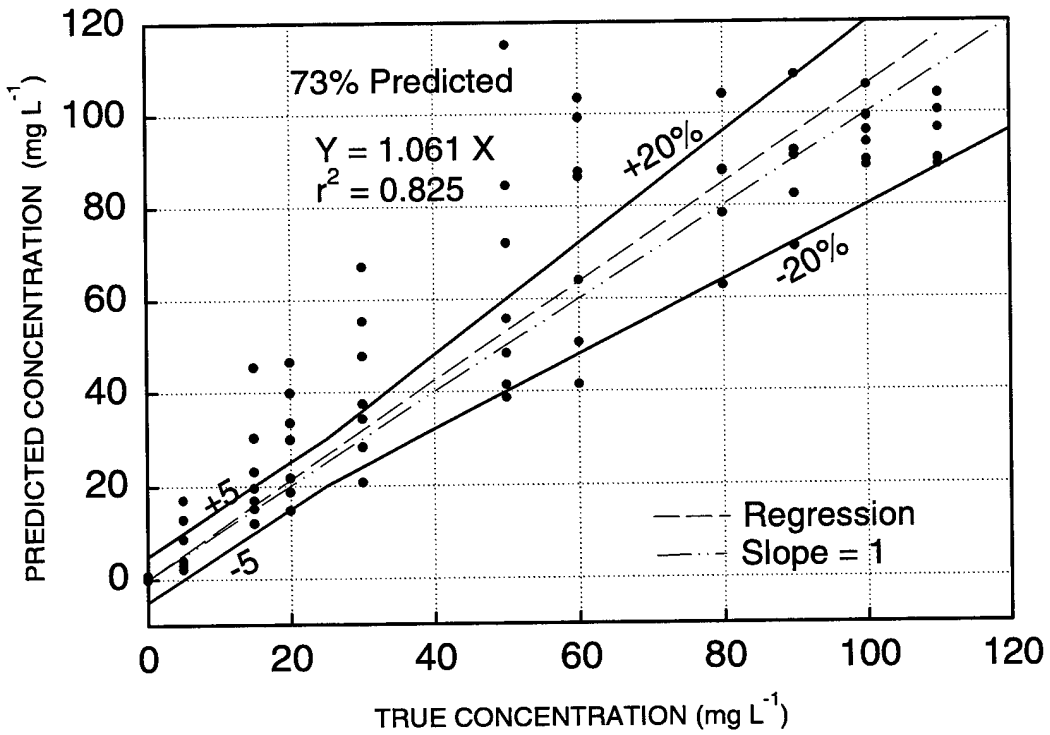

Figure 30. Prediction of oil concentrations for mixed oils with Tide® added. Network was trained with a combination of the data sets: single oils, mixed oils, and single oils added with detergents.

CONCLUSIONS

Both UV fluorescence and light scattering were used to characterize various oil samples from four oil types. Significant differences in fluorescence and scattering intensities were found not only among oil types, but also among separate samples, or subtypes within the same oil type. Furthermore, other factors such as oil mixing, salt content, and detergent concentration greatly affected the intensity of both fluorescence and scattering of oils. The interfering strength decreased as follows: Mil-D > Tide®> Oil type > Seawater > Mixing.

Before oil concentrations in wastewater can be measured accurately in real time, the analytical instrument must be calibrated. The traditional method—linear calibration— cannot be used for instrumental calibration. Using all four oil types in calibration results in a prediction rate of 28% (table 1). When the instrument is calibrated for each oil type with light scattering data, the prediction rates range from 20 to 67%. The all oil-type model with light scattering only predicts 20% of the oil samples (table 1). The use of multiple linear regression does not improve the prediction of the concentration of single oils.

A novel instrumental calibration method was developed for application to online measurement of oil concentrations in wastewater. The technique combines ultraviolet fluorescence and light scattering with an artificial neural network. Combining fluorescence and scattering greatly improves the prediction of oil concentrations compared to separate use of fluorescence or scattering —100% predicted for singles oils (table 1). The artificial neural network can predict the concentration of various oil types in the presence of seawater and detergents, with prediction rates ranging from 71 to 98% (table 2). With better defined training data sets, better prediction rates are expected for oil samples in the presence of chemical and physical interferences. The newly developed technique permits the online monitoring of oil spills, the accurate determination of oil concentrations in wastewater discharged from ships and the oil refinery industry, and oil detection in oil drilling fields.

REFERENCES

Andrews, J. M. and S. H. Lieberman. 1994. "Neural Network Approach to Qualitative Identification of Fuels and Oils from Laser Induced Fluorescence Spectra," *Analytica Chimica Acta*, vol. 285, pp. 237–246.

Andrews, J. M. and S. H. Lieberman. 1998. "Multispectral Fluorometric Sensor for Real Time In-Situ Detection Of Marine Petroleum Spills." *In Oil and Hydrocarbon Spills, Modeling, Analysis, and Control*, pp. 291–302, R. Garcia-Martinez and C. A. Brebbia, Eds. Computational Mechanics Publications, Boston, MA.

Lieberman, S. H. 1998. "Direct-Push, Fluorescence-Based Sensor Systems For *In Situ* Measurement of Petroleum Hydrocarbons in Soils," *Field Analytical Chemistry and Technology*, vol. 2, no. 2, pp. 63–73.

Mowery, R. L., H. D. Ladouceur, and A. Purdy. 1997. "Enhancement to the ET-35N Oil Content Monitor: A Model Based on Mie Scattering Theory." *NRL/MR/6176–97–7952*, Naval Research Laboratory, Washington, DC.

Nardella, J. A., T. A. Raw, and G. H. Stokes. 1989. "New Technology in Oil Content Monitors." *Naval Engineers Journal* (Mar), pp. 48–55.

Parker, H. and G. D. Pitt. 1986. *Pollution Control Instrumentation for Oil and Effluents.* Graham and Trotman, Inc., Gaithersburg, MD.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing: Explorations in the Micro-Structures Of Cognition*, pp. 318–362, D. E. Rumelhart and J. L. McCleland, Eds. MIT Press, Cambridge, MA.

Wasserman, P. D. 1989. *Neural Computing: Theory and Practice.* Van Nostrand Reinhold, New York, NY.

Zupan, J. and J. Gasteiger. 1991. "Neural Networks: A New Method For Solving Chemical Problems or Just A Passing Phase?" *Analytica Chimica Acta*, vol. 248, pp. 1–30.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 2000	Final

4. TITLE AND SUBTITLE

ONLINE MONITORING OF OILS IN WASTEWATER USING COMBINED ULTRAVIOLET FLUORESCENCE AND LIGHT SCATTERING WITH AN ARTIFICIAL NEURAL NETWORK

5. FUNDING NUMBERS

PE: 0601153N
AN: DN888573
WU: ME02

6. AUTHOR(S)

L. M. He
San Diego State University Foundation

L. L. Kear-Padilla
Computer Sciences Corporation

S. H. Lieberman
J. M. Andrews
SSC San Diego

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SSC San Diego
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 1816

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Environmental Quality
Technology Program
800 North Quincy Street
Arlington, VA 22217–5660

Naval Submarine Surface Warfare Center
Carderock Division
9500 MacArthur Blvd.
Bethesda, MD 20084–5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Ultraviolet (UV) fluorescence and light scattering are two analytical methods commonly used in instrumentation for online measurement of oils in water. UV fluorescence-based instruments detect both dissolved and emulsified aromatic constituents of oils. Light-scattering-based sensors measure optical scattering induced by emulsified oil droplets. A major technical challenge for each of these methods is to maintain quantitative accuracy in the presence of chemical and physical interferences, including fluorescent organic compounds (e.g., detergents and natural organic matter), suspended solid particles, dissolved salts, etc. To address this issue, we have been developing a new monitoring system that simultaneously combines both UV fluorescence and light scattering spectroscopy. Four major types of oils (lube oil 2190 and 9250, diesel fuel marine, and JP5), each of which had a dozen subtypes of oil samples, were examined to obtain the intensity of both fluorescence and scattering as a function of oil, detergent (Mil-D and Tide®), and seawater concentrations. Both fluorescence and light scattering intensities varied significantly with oil types and subtypes. Both Mil-D and Tide greatly influenced the fluorescence and scattering of oil samples.

14. SUBJECT TERMS

Mission Area: environmental chemistry/biotechnology
environmental programs
light scattering
ultraviolet fluorescence
online measure of oils in water

15. NUMBER OF PAGES

54

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(include Area Code)*	21c. OFFICE SYMBOL
S. H. Lieberman	(619) 553–2778 e-mail: lieberma@spawar.navy.mil	D361

The tremendous variations in fluorescence and scattering intensity with oil types and subtypes, detergents, and seawater make it difficult to calibrate the analytical instrument using traditional methods; hence we have implemented a multivariate, nonlinear calibration of instrumental response through an artificial neural network. We have demonstrated that the simultaneous, combined use of fluorescence and scattering data significantly improves quantitative prediction accuracy. The trained backpropagation neural network was used successfully to predict the concentrations of single oils and their mixtures, even in the presence of detergents and seawater, and appears well suited for calibrations of an online oil content monitor. The trained network processes information very quickly and is appropriate for real-time applications. The newly developed technique permits the online monitoring of oil spills, the accurate determination of oil concentrations in wastewater discharged from ships and the oil refinery industry, and oil detection in oil drilling fields.

INITIAL DISTRIBUTION

D0012	Patent Counsel	(1)
D0271	Archive/Stock	(6)
D0274	Library	(2)
D027	M. E. Cathcart	(1)
D0271	D. Richter	(1)
D361	S. H. Lieberman	(45)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (4)

SSC San Diego Liaison Office
Arlington, VA 22202–4804

Center for Naval Analyses
Alexandria, VA 22302–0268

Navy Acquisition, Research and
Development Information Center
Arlington, VA 22202–3734

Government–Industry Data Exchange
Program Operations Center
Corona, CA 91718–8000